Maximising the diagnostic value of structural MRI in the diagnosis of dementia

A comprehensive study of post-mortem proven cases

Submitted to University College London
for the degree of
Doctor of Philosophy

Lorna Harper
Dementia Research Centre
Institute of Neurology
University College London
September 2015

To my parents and gran.

This PhD is as much yours as it is mine.

Declaration

I, Lorna Harper, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis investigates the use of atrophy patterns from structural brain imaging to distinguish different dementia pathologies, including Alzheimer's disease, dementia with Lewy bodies and frontotemporal dementia pathologies. Using gold standard histopathology to stratify groups, analysis is based on 3D-T1-weighted imaging acquired during life in patients who attended clinic in one of three European centres. As well as comparison of disease groups with healthy controls, more clinically relevant comparisons between disease groups are performed to identify features that may be useful for differential diagnosis.

The image analysis techniques used in this thesis range from simple visual assessment to more advanced machine learning. Visual rating scales were found to be reliable, quick to perform, and when used in combination, could achieve diagnostic accuracy equal to unstructured visual assessment by dementia experts. Voxel based morphometry, used to provide a comprehensive estimate of global patterns of atrophy in pathologically distinct dementias, confirmed findings in the literature based on clinical data, and identified novel regions of interest for further study. A fully automated diagnostic approach using multi-atlas segmentation propagation and support vector classifiers, revealed brain volume differences between pathologically distinct groups, yet with several technical limitations to address.

Since histopathological diagnosis is rare in such a large, pathologically diverse cohort, this thesis also considers opportunities to develop the dataset into a shared resource for the dementia research community. To this end, a web application was developed to allow the data to be shared between collaborating centres, with plans to adapt this into a teaching resource.

In summary, this thesis uses a variety of analysis techniques to identify imaging features that may be useful for the differential diagnosis of dementia pathologies. Various opportunities are explored to maximise the value that can be derived from this unique and valuable dataset.

Acknowledgements

I would like to thank my supervisors Dr Jonathan Schott, Dr Ged Ridgway and Prof. Nick Fox for their help and advice throughout my PhD. I'm particularly grateful for their flexibility and support during my secondment.

I would also like to thank all the lovely, clever people at the Dementia Research Centre, past and present, for making it such a great place to work. In particular, I'm very grateful to Dr Kelvin Leung, Dr Ian Malone, Dr Casper Nielsen and Stuart Luscombe for pointing out various useful tools, particularly in the early development stages of this PhD, often answering questions I didn't know to ask. Dr Emma Burton (Newcastle University), and Dr Femke Bouwman and Dr Christiane Moller (VuMC, Amsterdam), collated the data at their respective centres before transferring to the Dementia Research Centre for processing. Dr Mark White (UCL) provided software for converting NIfTI image files into DICOM format. Dr Miklos Espak (UCL) and Dr Mark Austin (Ixico) assisted in the deployment of the web application. Dr Giorgio Fumagalli, Prof. Frederik Barkhof (VuMC, Amsterdam), Prof. Philip Scheltens (VuMC, Amsterdam), Prof. John O'Brien (Cambridge University), Prof. Nick Fox (UCL) and Dr Jonathan Schott (UCL) provided visual assessments and visual ratings for analysis. The Seeing Dementia application was developed in collaboration with Dr Ferran Prados Carrasco (UCL).

I would also like to give special thanks to the people with dementia and their families whose selfless dedication to research allowed me to carry out this work.

I'm forever grateful to Stuart, my parents, my gran, Kirsty and Dom for their constant support during the PhD, as always. Stuart put in a particularly hard shift and this thesis would not have been possible without him.

Contents

List of Figures

List of Tables

Thesis Aims

Distinguishing between the different neurodegenerative causes of dementia is vitally important to allow affected individuals and their families to access appropriate treatment, support and care [1], and this requirement will be even more pressing as/when disease-modifying therapies become available. Accurate diagnosis during life can be challenging as distinct underlying pathologies can result in overlapping clinical symptoms [2], and with the exception of rare autosomal dominant forms of dementia, histopathological diagnosis currently remains the gold standard [3]. However, in the context of thorough clinical examination certain patterns of atrophy, evident on structural MRI, have some predictive value in the differential diagnosis of neurodegenerative dementias. The evidence base is however limited in many cases, and relatively few studies have assessed atrophy patterns in the knowledge of underlying pathology. Therefore, using MR imaging acquired during life in a large multi-centre cohort of pathologically proven dementias, the aims of this thesis were:

1. To distil characteristic imaging features reported in the literature into a clinically useful algorithm that can be used to interrogate structural imaging, and to test its accuracy against pathological diagnosis.

2. To investigate the diagnostic value of visual rating scales widely used in dementia research, and the feasibility of applying them in a clinical setting.

3. To provide a comprehensive voxel-wise estimate of the global patterns of atrophy associated with different dementia pathologies, and to critically evaluate differential patterns that may help to guide visual assessment tools and region of interest analysis.

4. To assess regional differences in brain volume that can help to distinguish between different dementia pathologies.

5. To determine the additional diagnostic value of combining standard measures of assessment in an automated multivariate classifier.

6. To develop methods to share this unique and valuable dataset with the wider dementia research community.

Chapter 1

Introduction

Dementia is an umbrella term that describes progressive cognitive deficits, sufficient to impact on daily living, and is associated with a number of neurodegenerative disease processes, with Alzheimer's disease being the most common. In addition to the impact on the affected individual, these devastating diseases place a huge emotional and physical burden on caregivers, which intensifies as symptoms progress to include a loss of mobility, communication, and self care. Median survival is estimated to be around four and a half years, although may be ten years or more if disease onset occurs before 70 years of age [4].

Dementia is a global health problem. Current estimates suggest 5-8% of people over 60 are affected, equating to 47.5 million people worldwide, with this figure predicted to triple by 2050 to 135.5 million as the world's population ages [5]. In the UK alone, dementia is thought to affect 670,000 people (6.5% of the population) [6], costing the economy around £23 billion a year, with the majority of these costs currently absorbed by people with dementia and their families [7]. The potential socioeconomic impact of these diseases is, therefore, profound. In recognition of this fact, many of governments around the world have adopted long-term national dementia strategies [8–10] and international efforts have been consolidated through the launch of the Global Action Against Dementia programme, and the formation of the World Dementia Council at the G8 Dementia Summit, held in the UK in December 2013.

Now represented internationally at the governmental level, there are opportunities to influence global research spending and policy for the benefit of dementia research. Using this leverage should smooth the way for greater scientific discovery into the mechanisms driving these complex neurodegenerative diseases and ultimately lead to the discovery of disease modifying therapies. Even with such a breakthrough, however, it will remain important to help people live well with dementia, a process that starts with **early and accurate diagnosis** [11].

In this chapter, the diagnosis of dementia is discussed in terms of pathology, genetics and clinical features. Focussing on the three major neurodegenerative forms of dementia studied in this thesis, namely Alzheimer's disease, dementia with Lewy bodies (DLB) and frontotemporal lobar degeneration (FTLD), the challenges of accurate diagnosis are discussed in relation to the clinical tools currently available to extract diagnostically useful features.

1.1 Diagnosis of Dementia

Pathologically, the degenerative dementias are linked by protein misfolding in the brain, with the primary abnormal protein and its pattern of deposition currently used to define each neurodegenerative disease. Clinically, the dementias are less well defined, with distinct underlying pathology often producing overlapping clinical symptoms, and this is further complicated by the possibility of multiple comorbidities affecting the brain that are increasingly prevalent with advancing age. In this section, the current consensus guidelines for pathological diagnosis are reviewed, with reference to genetic causes and risks, followed by the associated criteria for clinical diagnosis during life. A list of all current diagnostic guidelines are included in Table 1.1.

1.1.1 Alzheimer's Disease

Pathology

Intraneuronal fibrils of primarily abnormal tau (neurofibrillary tangles (NFTs)), and extracellular deposits of β-amyloid (Aβ) peptides (senile plaques) are essential for a neuropathological diagnosis of Alzheimer's disease [12]. The mechanism which triggers these abnormal accumulations is widely speculated [13–17] but remains unproven, as indeed does the normal function and/or clearance of these proteins in the healthy brain [18–20]. The distinct progression of the accumulation of Aβ-plaques and NFTs across brain regions is, however, undisputed and continues to be used for staging the severity of disease [12, 21].

Based on the latest National Institute on Ageing and the Alzheimer's Association (NIA-AA) criteria, assessment of Alzheimer's disease neuropathologic change should incor-

porate:

A Thal phases of Aβ-plaque accumulation, which start in the neocortex, progressing to the hippocampal region, the diencephalon and striatum, brainstem and then the cerebellum [22].

B. Braak and Braak staging of NFTs, which progresses from transentorhinal involvement, to include limbic regions and extension into the neocortex [23, 24].

C. Scoring of neuritic plaques (a subset of senile plaques) according to the Consortium to Establish a Registry for Alzheimer's disease (CERAD) protocol, which is primarily based on frequency (sparse, moderate, frequent) rather than distribution [25].

To improve inter-rater reliability Aβ and NFTs scoring are reduced to four stages (see Figure 1.2). Based on the assessment of all three scoring systems, Alzheimer's disease neuropathologic change is classified as "Not", "Low", "Intermediate" or "High", with "Intermediate" or "High" considered a sufficient underlying cause of dementia (see Figure 1.2).

While Alzheimer's disease neuropathologic changes can exist in a pure form, co-morbidities are increasingly prevalent with advancing age [27]. The NIA-AA, therefore, also recommends assessment of the most common of these co-morbidities during post mortem examination, which include: Lewy body disease (included in Figure 1.2), vascular brain injury, hippocampal sclerosis, argyrophilic grain disease and transactive response DNA binding protein 43 kDa (TDP43) inclusions [12, 21]. These pathologies, which may also occur in a pure form, are discussed as such in subsequent sections.

Genetics

Fully penetrant, autosomal dominant mutations associated with Alzheimer's disease occur on one of three genes: the *amyloid precursor protein (APP)* gene [28], the *Presenilin 1 (PSEN1)* gene [29] or the *Presenilin 2 (PSEN2)* gene [30]. Pathogenic mutations in the *APP* gene result in increased production of Aβ and/or alteration in the ratio of specific Aβ peptides, and are associated with cerebral amyloid angiopathy (CAA) as

A Score	Thal phase for Aβ-plaques	B Score	Braak and Braak NFT stages	C Score	CERAD neuritic plaque score
0	None	0	None	0	None
1	1 or 2	1	1 or 2	1	Sparse
2	3	2	3 or 4	2	Moderate
3	4 or 5	3	5 or 6	3	Frequent

Figure 1.1: NIA-AA scoring of Alzheimer's disease neuropathologic change. Three established pathology rating scales (Thal phases, Braak and Braak and CERAD) are mapped onto three four-point scales (A, B, C respectively). CERAD=Consortium to Establish a Registry for Alzheimer's Disease, NFTs=Neurofibrillary tangles.

A	C	B0 or B1	B2	B3	Lewy bodies
0	0	Not [1]	Not[1]	Not [1]	None
1	0 or 1 2 or 3[3]	Low Low	Low Int. AD	Low[2] Int. AD[2]	Brainstem
2	Any C	Low AD[4] (High DLB)	Int. AD (Int. DLB)	Int. AD[2]	Limbic
3	0 or 1 2 or 3	Low AD[4] (High DLB) Low AD[4] (High DLB)	Int. AD (High DLB) Int. AD (High DLB)	Int. AD[2] (Int. DLB) High AD (Int. DLB)	Neocortical

Figure 1.2: NIA-AA Criteria for Alzheimer's disease neuropathologic change based on the ABC scores outlined in Figure 1.1 (adjusted for additional Lewy body disease inclusions (see Section 1.1.2)). Liklihood of pathological Lewy body disease is assessed in relation to Braak and Braak staging of neurofibrillary tangles according the McKeith criteria [26]. AD=Alzheimer's disease, DLB=dementia with Lewy bodies, Int.=Intermediate, NIA-AA=National Institute on Ageing and the Alzheimer's Association

1. Medial temporal lobe NFTs in the absence of significant Aβ or neuritic plaques occurs in older people and may be seen in individuals without cognitive impairment, with mild impairment, or with cognitive impairment from causes other than Alzheimer's disease.
2. Widespread NFTs with some Aβ/amyloid plaques but limited neuritic plaques is relatively infrequent and when it occurs, other diseases, particularly tauopathies, should be considered.
3. High levels of neuritic plaques in setting of low Thal phase is a rare occurrence and should prompt reconsideration of neuritic versus diffuse plaques, and the possible contribution of other diseases to cognitive impairment or dementia.
4. Higher levels of Aβ or neuritic plaques with low Braak stage should prompt consideration of contribution by co-morbidities like vascular brain injury, Lewy body disease, or hippocampal sclerosis.

well as Alzheimer's disease [31]. The proteins encoded by the *PSEN1* and *PSEN2* genes are directly involved in the processing of *APP*, and mutations in these genes also lead to an increase in Aβ production [31]. *PSEN1* mutations are the most common cause of familial Alzheimer's disease, while *PSEN2* mutations are very rare. Familial Alzheimer's disease accounts for around 0.5% of all Alzheimer's disease cases (http://www.molgen.ua.ac.be/ADMutations/).

The more common "sporadic" form of Alzheimer's disease is influenced by a number of risk factor genes. To date, the *APOE* $\varepsilon 4$ allele confers the greatest risk from a single gene, and is known to be dose dependent, with increased risk associated with $\varepsilon 4$ homozygotes [32]. More recently, a *TREM2* (triggering receptor expressed on myeloid cells 2) heterozygous mutation was also identified as a moderate risk for Alzheimer's disease [33], while advances in sequencing technologies have revealed several additional genetic risk factors, each conveying a relatively small risk for developing the disease [31, 34]. Importantly, the identification of these genes continues to shed light on the biological pathways of the disease and provides opportunities to investigate alternative therapeutic approaches [35].

Clinical Presentation

There is an increasing interest in identifying presymptomatic forms of Alzheimer's disease [36, 37], and patients with mild cognitive problems due to Alzheimer's disease, termed "mild cognitive impairment" [37, 38]. For most patients, however, a clinical diagnosis of Alzheimer's disease starts with evidence of dementia, to include cognitive and behavioural symptoms that 1) interfere with the ability to function in daily life, 2) represent a decline from previous levels of functioning, and 3) are not explained by delirium or a major psychiatric disorder [39]. According to the most recent NIA-AA guidelines, assessment should subsequently include testing for a number of clinical features and biomarkers (discussed in Section 1.2) that add a degree of certainty to the diagnosis, i.e. possible or probable Alzheimer's disease [39]. In addition to the typical amnestic presentation of Alzheimer's disease, the guidelines acknowledge atypical variants of Alzheimer's disease, which include logopaenic aphasia (language variant), posterior cortical atrophy (visual variant), and a frontal, dysexecutive variant. These atypical classifications are also included in the most recent recommendations

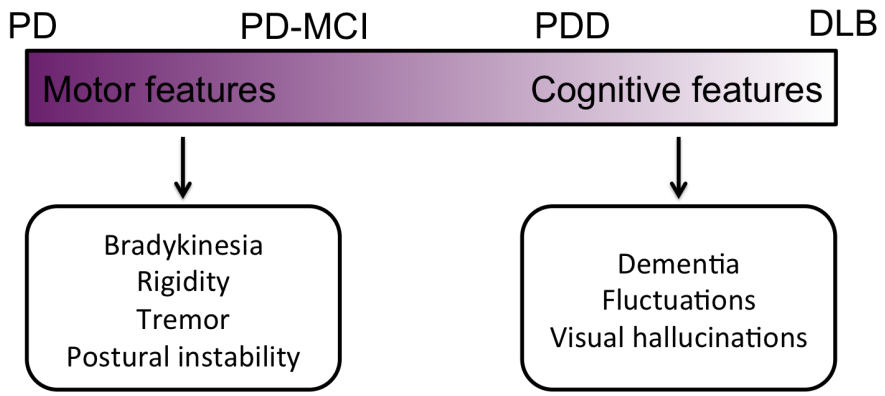


Figure 1.3: Spectrum of clinical symptoms associated with Lewy body disease. PD=Parkinson's disease, PD-MCI=Parkinson's disease - mild cognitive impairment, PDD=Parkinson's disease dementia, DLB=Dementia with Lewy bodies

from the International Working Group (IWG-2) research criteria for the diagnosis of Alzheimer's disease [37].

1.1.2 Dementia with Lewy Bodies

Pathology

Lewy bodies and Lewy neurites are pathological aggregates containing α-synuclein, and are the core features required for the histopathological diagnosis of a range of Lewy body diseases, which includes Parkinson's disease and DLB [26]. Lewy body disease produces a spectrum of clinical symptoms, from primary motor features (pure Parkinson's disease) to primary cognitive impairment (pure DLB) (see Figure 1.3)[40]. Phenotypic manifestations within this spectrum (i.e. motor to cognitive) has been shown to correlate with the distribution of α-synuclein, from brainstem nuclei, to limbic, and neocortical regions [41, 42], although this is not true in all cases and the disease mechanisms that lead to dementia are likely to be complex [43]. Lewy body disease is however, frequently found in the presence of Alzheimer's disease neuropathological change, therefore, at the cognitive end of the spectrum, a pathological diagnosis of DLB requires both evidence of Lewy body pathology and a relative absence of coexisting Alzheimer's disease pathology (see Figure 1.2) [21, 26].

Genetics

Very few cases of familial DLB have been described, a causal gene has not been identified, and DLB is generally considered a sporadic disease [44]. There are however, known genetic risk factors for DLB (e.g. GBA) [45], some of which are shared with Alzheimer's disease (e.g. *APOE* $\varepsilon 4$) and Parkinson's disease (e.g. SNCA, SCARB2) [46].

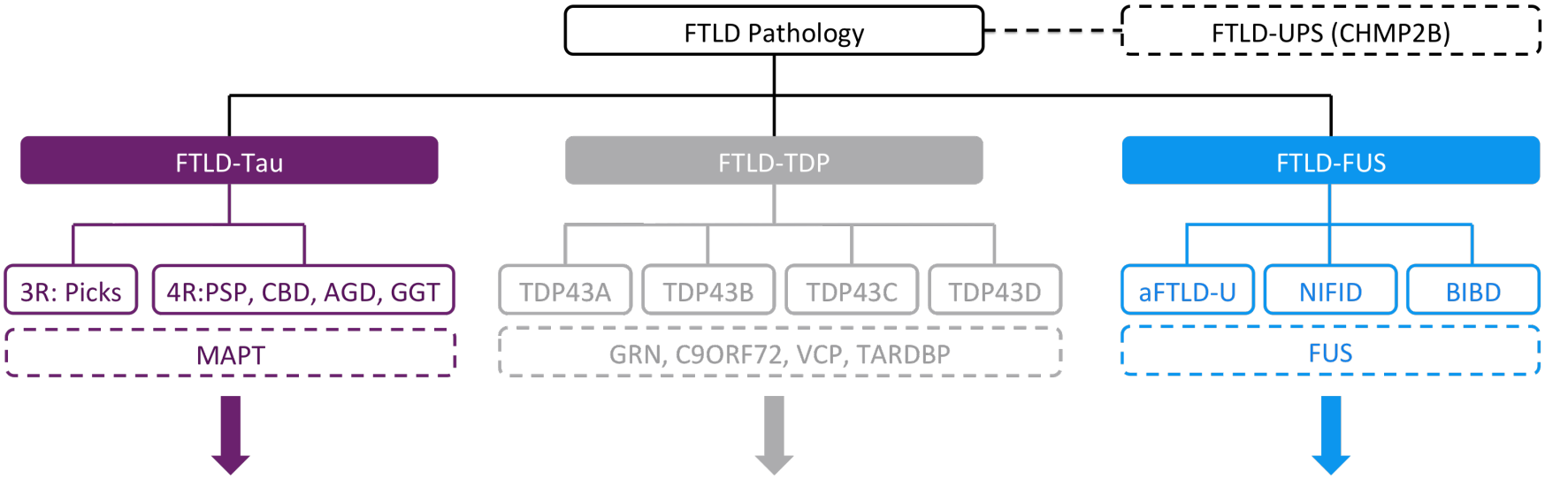
Clinical Presentation

Whilst milder forms of α-synuclein related cognitive impairment are recognised, dementia is a core feature essential for the diagnosis of DLB. A typical presentation includes deficits in attention, executive function and visuospatial ability, with relatively preserved memory early in the disease course [26]. Additional features including fluctuating cognition, recurrent visual hallucinations and spontaneous features of parkinsonism, add further support to the diagnosis (see consensus criteria for details [26]). The temporal sequence of symptoms is also important for the diagnosis of DLB and its differentiation from Parkinson's disease dementia. DLB is diagnosed when dementia occurs before or concurrently with parkinsonism (within one year), while Parkinson's disease dementia should be diagnosed if dementia develops in the context of established Parkinson's disease [26]. Biomarker techniques used for the diagnosis of DLB are discussed in Section 1.2.

1.1.3 Frontotemporal Lobar Degeneration

Pathology

Whilst there is considerable clinical overlap, four major pathological proteinopathies underlie FTLD. FTLD-Tau and FTLD-TDP43 account for the majority of FTLD pathology (~40% respectively) [47], while around 10-15% are immunoreactive for fused in sarcoma protein (FUS) [48]. A rare fourth class, only detectable with immunohistochemistry against proteins of the ubiquitin proteasome system (FTLD-UPS), is used to describe the pathology associated with rare CHMP2B mutations [48] (see Figure 1.4).



Tauopathy	Associated clinical syndromes
Picks	bvFTD, PNFA, SD
PSP	PSPS, CBS, PNFA
CBD	CBS, PSPS, bvFTD, PNFA
AGD	bvFTD
GGT	bvFTD, FTD-MND, CBS, PSPS
MAPT	bvFTD, PSPS, CBS

TDP43opathy	Associated clinical syndromes
A (GRN, C9ORF72)	bvFTD, FTD-MND, PNFA, CBS
B (C9ORF72)	bvFTD, FTD-MND, PNFA
C	SD, bvFTD
D (VCP)	bvFTD
TARDBP	MND

FUSopathy	Associated clinical syndromes
aFTLD-U	bvFTD
NIFID	bvFTD, PSPS, CBS
BIBD	FTD-MND, MND, PSP
FUS	MND

Figure 1.4: FTLD pathology stratified by molecular class [48]. Genetic mutations associated with these pathologies are listed in the dashed boxes. Associated clinical phenotypes are listed in the tables below [50, 56, 57]. FTLD-UPS is used to describe the pathology associated with rare CHMP2B mutations.

Glossary: 3R: 3 repeat tau, 4R: 4 repeat tau, AGD: Agyrophilic grain disease, aFTLD-U: Atypical FTLD with ubiquitin inclusions, BIBID: Basophilic inclusion body disease, bvFTD: behavioral variant FTD, C9ORF72: Chromosome 9 open reading frame 72, CBD: Corticobasal degeneration, CBS: Corticobasal syndrome, FUS: Fused in sarcoma protein, GGT: Globular glial tauopathies, NIFID: Neuronal intermediate filament inclusion disease, MND: Motor neuron disease, Picks: Pick's disease, PNFA: Progressive non-fluent aphasia, PSP: Progressive supranuclear palsy, PSPS: Progressive supranuclear palsy syndrome, SD: Semantic dementia, TARDBP: TAR DNA-binding protein 43 (mutation), VCP: Valosin-containing protein.

FTLD-Tau is characterised by lesions containing insoluble tau protein, which can be further classified by assessment of the predominant isoforms of tau present in these lesions (for more info see Appendix A) and their regional distribution within the brain [49]. Unlike the NFTs found in Alzheimer's disease which contain all six isoforms of tau, the majority of lesions associated with FTLD-Tau primarily contain either 3-repeat or 4-repeat tau [49].

TDP43 binds to both DNA and RNA and is integral to the production of many proteins, particularly those involved with neuronal development [50]. The pathogenesis of FTLD-TDP43 proteinopathies is therefore, complex, with evidence to suggest that toxic gains and/or functional losses may be associated with TDP43 dysregulation [51]. However, as a consistent pathological feature of several different genetic forms of FTLD (see below), dysregulation of TDP43 may be a downstream mechanism common to several disease pathways [50], including motor neuron disease (MND) and to a lesser extent Alzheimer's disease and DLB [52–54]. FTLD-TDP43 pathology can be further classified into types A, B, C and D based on the morphology and regional distribution of TDP43 inclusions, with ordering determined by the relative frequency of this pathology within the population (type A being the most common) [55].

FUS belongs to the FET (previously TET) protein family, which function in several aspects of cell growth control and therefore, like TDP43, its role in neurodegeneration is not yet fully understood [50]. However, the accumulation of FUS protein is the most prominent molecular pathology shared by the distinct clinicopathological entities grouped under FTLD-FUS, namely atypical FTLD with ubiquitin inclusions, neuronal intermediate filament inclusion disease and basophilic inclusion body disease [48].

Genetics

FTLD has a strong genetic component, with evidence of family history in around 40% of cases, typically with an autosomal dominant pattern of inheritance [58]. The frequency of occurrence varies by geography but the most common mutations occur on the *microtubule associated protein tau (MAPT)*, the *progranulin (GRN)*, and in the region of the *chromosome 9 open reading frame 72 (C9ORF72)*, with the latter also a

major genetic cause of MND. Less common mutations include the valosin-containing protein (*VCP*), charged multivesicular body protein 2B gene (*CHMP2B*), FUS (also associated with MND), and TDP43 (or *TARDP*).

Clinical Presentation

Frontotemporal dementia is the term used to describe the heterogeneous clinical syndromes that manifest in patients with underlying FTLD pathology. The most common clinical presentations are behavioural variant frontotemporal dementia (bvFTD) and language variant frontotemporal dementia, known as primary progressive aphasia. However, the FTLD spectrum is wide, encompassing cross-over with MND and parkinsonism. There is rarely a one-to-one mapping between frontotemporal dementia phenotypes and molecular pathology (or genetic mutation); however, some syndromes are found to be more frequently associated with certain pathologies as shown in Figure 1.4.

Criteria for bvFTD [59] requires progressive deterioration of behaviour and/or cognition with at least three of the following symptoms within three years of disease onset: behavioural disinhibition; apathy; loss of sympathy or empathy, preservative, stereotyped or compulsive/ritualistic behaviour; hyperorality and dietary changes; and a neuropsychological profile exhibiting executive deficits with relative sparing of memory and visuospatial functions. Additional evidence of significant functional decline and corroboration from brain imaging (discussed in Section 1.2) adds greater certainty to the diagnosis, from possible to probable bvFTD.

Primary progressive aphasia can be further subdivided into three specific categories with specific diagnostic criteria: the non-fluent variant, the semantic variant, and the logopaenic variant, however, the later is typically associated with Alzheimer's disease pathological changes [60]. More details on the clinical features associated with these categories is provided by Gorno-Tempini *et al* [60]. Warren *et al* also provide a simple algorithm for bedside clinical assessment [61].

Parkinsonian syndromes with primary motor features, such as corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP), which are the result of accumulation of abnormal 4-repeat-tau, may also exhibit frontal executive impairment or non-fluent

aphasia, and cognitive impairment is considered a supportive feature in their diagnostic criteria [62, 63]. MND-type motor disfunction with frontotemporal dementia features (FTD-MND) is also a recognised clinical phenotype associated with TDP43 or FUS pathology. Indeed, due to the extent of the overlap between MND and frontotemporal dementia at the clinical, pathological and genetic levels it has been proposed that these are not distinct disorders and should be considered two ends of a disease continuum [64].

1.2 Diagnostic Tools

With the exception of rare autosomal dominant forms of dementia, definitive biomarkers of neurodegenerative pathology are lacking, and accurate diagnosis during life relies on the interpretation and synthesis of evidence from multiple sources. In this section the diagnostic tools available to extract this evidence are reviewed.

1.2.1 Clinical Assessment

The foundation of an accurate diagnosis is a detailed clinical examination, and any additional biomarkers should only be interpreted in the context of these findings. Investigation starts with history taking from the patient and a knowledgeable informant, to include family history, vascular risk factors, and a review of medication. This initial investigation is perhaps the most important and an experienced clinician can determine a great deal from this initial interview. An objective cognitive assessment follows and typically includes tests of orientation, attention, memory, language, executive function, praxis and visuospatial skills, taking into account prior level of functioning and attainment. In cases of mild or uncertain dementia, formal neuropsychological testing can be useful to further probe cognitive deficits. Physical examination of the neurological system is used to detect sensory, motor and coordination deficits that may support neurodegenerative dementia or indicate other rarer and possibly treatable causes. Blood should be tested to exclude partially treatable causes of dementia, these include test of thyroid function, vitamin B_{12} levels, and infection if clinically indicated [65–67].

1.2.2 Neuroimaging

Neuroimaging can provide valuable diagnostic information about both the structure and function of the brain, which can be used to exclude certain non-degenerative pathologies or provide evidence to support subtype dementia diagnosis (see Figure 1.5). Table 1.1 provides a summary of imaging recommendations included in current diagnostic guidelines and commonly available imaging techniques are reviewed in this section, while molecular imaging is discussed in the following (Section 1.2.3).

Structural Brain Imaging

CT and MRI are the primary imaging modalities available for assessing brain structure. While much can be determined about brain structure from CT, MRI produces non-ionising radiation, allows for multiple aspects of brain tissue pathology to be explored, and is the structural imaging modality of choice in the assessment of dementia. 3D T1-weighted sequences provide good spatial resolution, high contrast between grey and white matter, and can be reformatted in three planes for thorough assessment of cerebral atrophy or the detection of other lesions that could be contributing to the patient's symptoms [68]. The diagnostic value of patterns of atrophy, detected on T1-weighted imaging, is the focus of this thesis. Methods to detect and interrogate these patterns are discussed in detail in Chapter 2, while characteristic imaging features previously described in the literature are discussed in Chapter 3.

T2-weighted and fluid attenuation inversion recovery (FLAIR) are used to detect vascular damage, but can also indicate inflammatory, metabolic, toxic or infective processes that could be contributing to cognitive deficits [69]. T2*-weighted imaging (or preferably susceptibility weighted imaging [70]) is used to detect cerebral microbleeds often associated with CAA, which may be important in the aetiology of Alzheimer's disease [71]. Diffusion weighted imaging provides a volume-averaged measure of the diffusion of water molecules within brain tissue, which is altered in various pathological states and is particularly useful for detecting recent infarct [72], or spongiform change associated with rare cases of Creutzfeldt Jakob disease [73]. For completeness, vascular features are also discussed in more detail in Chapter 3, however, they do not feature in the experimental work of this thesis.

	Pathology Guidelines	Clinical Guidelines	Imaging Recommendations
AD	NIA-AA: Hyman *et al* (2012) (Montine *et al* (2012))	NIA-AA: McKhann *et al* (2011) IWG-2: Dubois *et al* (2014)	**- Extensive infarcts or WMHs for a diagnosis of mixed-AD** **- Abnormal dopamine transporter PET scan for a diagnosis of mixed-AD** **- Decreased ^{18}F-FDG uptake or disproportionate atrophy in temporoparietal regions** **- Extensive infarcts or WMHs for a diagnosis of AD/atypical AD**
DLB	DLB-C: McKeith *et al* (2005)	DLB-C: McKeith *et al* (2005)	**- Low striatal DAT uptake on SPECT/PET** **- Relative preservation of MTL on CT/MRI** **- Generalised low uptake on SPECT/PET perfusion scan with reduced occipital activity** **- CVD evident on brain imaging**
FTLD	FTLD-C: Cairns *et al* (2007) ***Updates:*** Mackenzie *et al* (2009): TDP43 Mackenzie *et al* (2010): FUS Mackenzie *et al* (2011): TDP[ABCD]	bvFTD-C: Rascovsky *et al* (2011)	**Hypo-perfusion/metabolism or atrophy in frontal and/or anterior temporal lobe regions**
		PPA-C: Gorno-Tempini *et al* (2011)	**Hypo-perfusion/metabolism or atrophy in the:** **- left posterior fronto-insular region (PNFA)** **- anterior temporal lobe region (SD)** **- left posterior perisylvian or parietal (LPA)**
		CBD-C: Armstrong *et al* (2013)	**Structural lesion suggestive of focal cause**
		NINDS-SPSP: Litvan *et al* (1996) NNIPPS: Bensimon *et al* (2009)	**Relevant structural abnormality (i.e. basal ganglia/brainstem infarcts, lobar atrophy)**

Table 1.1: Current guidelines for pathological diagnosis and clinical diagnosis with imaging recommendations. Imaging recommendations are coloured based on the type of evidence they provide - **Core feature**, **Supportive feature**, **Exclusion criterion**. AD=Alzheimer's disease, DLB=dementia with Lewy bodies, FTLD=frontotemporal lobar degeneration, bvFTD=behavioural variant FTD, PPA=primary progressive aphasia, CBD=corticobasal degeneration, NIA-AA=National Institute on Ageing and the Alzheimer's Association, NINDS-SPSP=Neurological Disorders and Stroke and Society for Progressive Supranuclear Palsy, NNIPPS=Neuroprotection and Natural History in Parkinson Plus Syndromes

A basic MRI protocol consisting of standard sequences required for thorough assessment of relevant imaging features is presented in Figure 1.6. While thinner slices, and where possible 3D rather than 2D acquisitions, provide better spatial resolution to detect smaller lesions, there is a compromise to be made between the level of detail available for assessment and the time taken to acquire the image.

Functional Brain Imaging

Cerebral perfusion imaging with ^{99m}Tc-HMPAO single photon emission computed tomography (SPECT) is an established technique often used in the diagnosis of dementia. ^{99m}Tc-HMPAO is a lipophilic compound which passes through the blood brain barrier and undergoes hydrophilic conversion, quickly trapping the radiotracer in brain tissue (stabilising 10-15 minutes post injection) [74]. SPECT imaging provides an estimate of relative regional flow differences and the detection of regions of hypoperfusion associated with reduced neuronal activity [74]. ^{18}F-FDG is a glucose analog and therfore, provides a more direct measure of neuronal activity than ^{99m}Tc-HMPAO. Comparisons between perfusion imaging (SPECT) and metabolic imaging (positron emission tomography (PET)) suggest that while there may be equivalence in detecting people with dementia from controls, PET is superior in the differential diagnosis of Alzheimer's disease and DLB [75].

Different aspects of the pre-synaptic dopamine system, implicated in Parkinson's disease and other parkinsonian disorders (DLB, PSP, corticobasal degeneration (CBD), multiple system atrophy), can also be imaged in vivo using ^{123}I-FP-CIT SPECT or ^{18}F-FDOPA PET. While ^{123}I-FP-CIT labels dopamine transporters (see Figure 1.5), ^{18}F-FDOPA reflects the synthesis and storage of L-DOPA (a precursor of dopamine) in pre-synaptic vesicles. Direct comparison of both tracers has demonstrated equal sensitivity [76], however, there is some evidence to suggest a compensatory mechanism in the dopamine system causes upregulation of L-DOPA synthesis in the early stages of disease, potentially reducing the sensitivity of ^{18}F-FDOPA as an early biomarker [77].

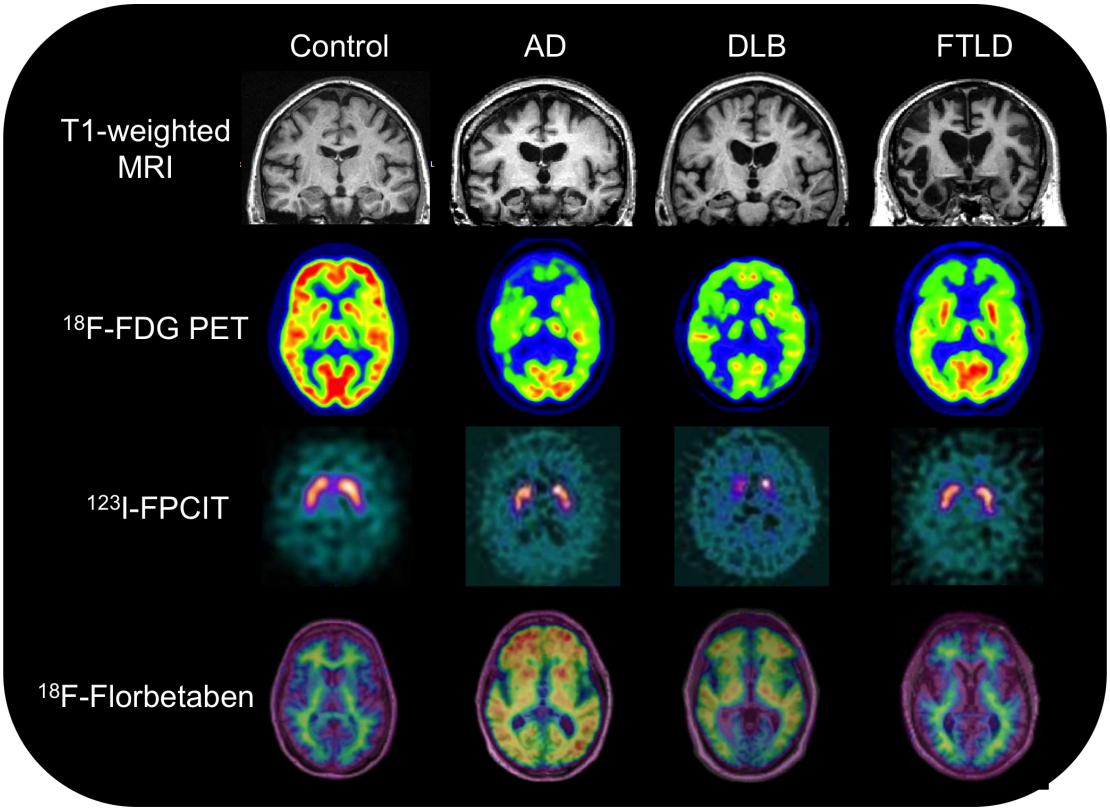


Figure 1.5: Neuroimaging in the diagnosis of dementia. MRI images are from pathologically confirmed cases. FDG PET images from Koeppe *et al* [78]. FPCIT images from Morgan *et al* [79] with pathological confirmation of Alzheimer's disease and DLB pathology. Only the DLB image is abnormal. Florbetaben images are overlaid on MRI images from Villemagne *et al* [80]. Nonspecific binding is shown in white matter in controls and FTLD, specific binding is shown in cortical areas of Alzheimer's disease and DLB patients.

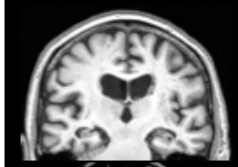
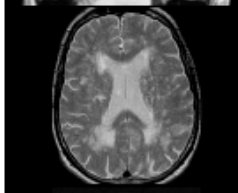
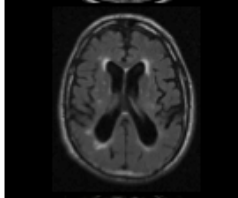
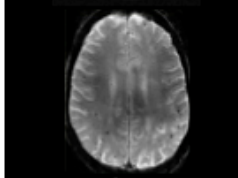
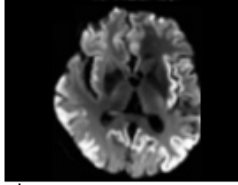
	Sequence	Orientation	Slice Thickness	Duration	Use
	3D T1-weighted	Coronal	1mm isotropic	8 mins	Detecting structural abnormalities and atrophy
	T2-weighted	Transverse	3-5 mm	4 mins	Detecting subcortical signal change
	FLAIR	Transverse	3-5 mm	4 mins	Detecting subcortical signal change
	T2* GE	Transverse	3-5 mm	4 mins	Detecting microbleeds
	DWI (ADC map)	Transverse	5 mm	2 mins	Detecting recent infarcts, spongiform change

Figure 1.6: Basic MRI protocol required for the clinical assessment of cognitive impairment. GE = gradient echo.

1.2.3 Molecular Biomarkers

Biomarkers of amyloid and tau pathology are now available, with the ability to identify some aspects of Alzheimer's disease pathology in earlier stages of the disease. Due to the technical requirements of these techniques they are not yet widely available. Their current use is discussed briefly below.

Cerebrospinal Fluid

cerebrospinal fluid (CSF) obtained by lumbar puncture provides a means of measuring biochemical changes that occur in the brain's extracellular space. Basic CSF biomarkers, such as cell counts, serum albumin and oligoclonal bands, can be be used to exclude infections, damage to the blood brain barrier, and inflammatory disorders which may mimic or contribute to dementia [81]. More recently, however, CSF biomarker techniques have been developed to measure core Alzheimer's disease pathology such as $A\beta_{1-42}$, phosphorylated tau (pTau) and total tau (tTau). In Alzheimer's disease $A\beta_{1-42}$ is low, while tTau (a more general marker of neuronal degeneration) and pTau (thought to be more specific for Alzheimer's disease) are both high. Current evidence suggests a combination of low $A\beta_{1-42}$ and high pTau or tTau provides optimal diagnostic specificity [82]. Establishing reference ranges and cut-points for these measures is challenging in the absence of confirmed pathology, and further complicated by a lack of standardisation in CSF collection and analysis [83, 84]. However, work is under way to address these issues and CSF analysis has recently been included in the research diagnostic guidelines for Alzheimer's disease [37].

Amyloid PET Imaging

While CSF provides a measurement of soluble $A\beta_{1-42}$, PET ligands, with their design based on histopathological stains such as congo red and thioflavin-T, provide a measure of aggregated fibrillar $A\beta$ [85]. Studies have demonstrated that both measures are well correlated [86], however, imaging provides the additional benefit of being able to localise amyloid deposition within brain tissue. Like CSF, standardisation of quantitative image analysis is still ongoing [87], however, simple visual reads of amyloid 'positive' or 'negative' are recommended for diagnostic use [37] and have been validated

against post mortem assessment of Alzheimer's disease pathology [88] (see Figure 1.5). There are currently four ^{18}F amyloid PET tracers licensed for use in the diagnosis of Alzheimer's disease: ^{18}F-Florbetapir/AMYViD (Avid) [89], ^{18}F-Flutemetamol/Vizamyl (GE) [90], ^{18}F-Florbetaben/Neuraceq (BV Cyclotron VU) [91], ^{18}F-Flutafuranol (AstraZeneca/Navidea) [92]. Despite different grey matter and white matter retention characteristics, preliminary studies suggest there is no clear advantage to using one over the other [93].

Tau Imaging

Tau imaging is the next major diagnostic milestone in dementia and has the potential to provide greater insights into the development and treatment of major tauopathies, such as Alzheimer's disease and some variants of FTLD [94]. In Alzheimer's disease research in particular, it may help to answer some long standing questions relating to the role of tau pathology in the degenerative process and its relation to Aβ deposition [94, 95]. Due to the complexities of tau aggregation and the requirements for tracer design (for a review see Villemagne *et al* [95]), the development and validation of tau tracers is complex and still ongoing. Significant advances have been made, moving from non-selective tau tracers, which also bind to extracellular Aβ plaques [96, 97], to tracers with more selective binding affinity for tau pathology (reviewed by Villemagne *et al* [95]), however, questions still remain over tracer kinetics and some unexplained focal retention inconsistent with neuropathological reports [95]. Tau imaging is still in the development and testing stage and is a long way from being used for clinical diagnosis, however, it has great potential to advance current understanding of neurodegenerative disease processes and evaluate prospective therapies.

1.3 Conclusion

The neurodegenerative dementias are complex diseases and our understanding of their aetiology is limited. Histopathology allows for the stratification of disease by primary abnormal molecular component but in many cases it remains unclear if these are the downstream effects of a more complex disease pathway. Genetics continues to provide the greatest insights into underlying disease mechanisms and increased understanding

of genetic risk factors and environmental interactions may hold the ultimate key to therapeutics and improved diagnostics.

In-depth study of early cognitive features and symptom progression have led to greater appreciation of distinct clinical phenotypes and this information has been incorporated into the latest diagnostic guidelines. While there is currently no one-to-one mapping between the clinical features appearing during life and the pathology evident at post mortem, in many cases these features are more indicative of certain pathologies and their identification increases the probability of accurate clinical diagnosis.

Biomarkers continue to evolve, helping to further refine clinical phenotypes in some cases and provide direct evidence for aspects of the underlying pathology. However, many of these techniques are highly complex, expensive and remain unavailable outside of specialist centres. In the absence of disease modifying therapies, this is unlikely to change as healthcare systems struggle financially to cope with the demands of an ageing population. In the meantime, the existing international commitment to help people live well with dementia requires optimising the use of the diagnostic techniques that are more widely available, such as MRI.

Structural brain imaging, particularly MRI, is not only widely available, it is also multi-purpose, and continues to be recommended for use in all diagnostic guidelines. Firstly, it can be used to exclude surgically treatable mass-effect lesions that, due to their size and or location within the brain, could be the source of cognitive impairment. Secondly, it can be used to characterise and determine the extent of vascular lesions, which are increasingly recognised as an important contributor to cognitive decline. Although the role of vascular disease is not fully understood, vitally for patients it can be managed and treated, and its assessment is recommended in all diagnostic guidelines, forming part of the exclusion criteria for pure Alzheimer's disease and DLB (see Table 1.1). Finally, MRI can also be used to detect patterns of cerebral volume loss with predictive value for underlying dementia pathology. While no pattern is 100% sensitive or specific, certain characteristic features (described in Chapter 3) can provide good supporting evidence for the diagnosis of specific dementia subtypes.

Cerebral volume loss is recognised as a downstream effect of several neurodegenerative

disease processes [98], yet changes can be detected in early [99] and even presymptomatic [100] stages of disease. It is well established as an outcome measure in clinical trials [101], and is included as a core or supporting feature in most diagnostic guidelines [36, 38, 39, 59, 60, 65]. While there is an extensive evidence base to support the use of cerebral volume loss in the diagnosis of dementia, much of this evidence is derived from the study of clinically diagnosed dementias, which even in specialist centres, may be inaccurate in around 30% of cases [3]. Identifying patterns of volume loss that help to separate pathological processes from normal ageing, and from one another, based on gold standard histopathology diagnosis, may help to translate these imaging findings from the research domain to routine clinic practice. This thesis aims to address this issue using a rarely-available, large, pathologically-proven dataset with standard 3D T1-weighted imaging acquired in patients with the three primary neurodegenerative dementias (Alzheimer's disease, DLB and FTLD). Using a variety of image analysis techniques from basic visual assessment to more computationally advanced classification techniques, this thesis investigates the opportunities to maximise the value of routine structural MRI in the diagnosis of dementia.

Chapter 2

Techniques for Extracting Diagnostic Value from Structural MRI

2.1 Introduction

As discussed in Chapter 1, structural MRI is widely available and routinely acquired in the assessment of cognitive impairment. Characteristic patterns of atrophy, discussed in detail in Chapter 3, are known to have predictive value in the diagnosis of dementia but accurate and reliable methods of quantifying and extracting this information are required in order to make use of it for clinical or research purposes. In this chapter, a variety of analysis techniques based on the assessment of atrophy are discussed, from simple visual assessment to more complex automated analysis. Current evidence for their use in neurodegenerative dementia is discussed in Chapter 3. As the focus of this thesis is on the diagnostic value of structural MRI, discussion will centre around cross-sectional imaging techniques, however, longitudinal imaging also provides valuable insights into disease progression [102, 103] and is often used as an outcome measure in clinical trials [104, 105].

2.2 Qualitative Assessment

Qualitative assessment of structural imaging does not require any specialist software or computing capabilities, and can be performed on any standard radiology workstation, making it suitable for implementation into clinical practice. It does, however, require expert knowledge of an array of imaging features and patterns of atrophy associated with the neurodegenerative dementias. In this section, we briefly discuss some of the features that can be detected from basic visual assessment (a more in depth discussion of these features can be found in Chapter 3) as well as the tools available to help struc-

ture image analysis, making the diagnostic information it contains more accessible to a wider clinical audience.

2.2.1 Basic Visual Assessment

Unstructured visual assessment of brain imaging in the diagnosis of dementia is the primary means of analysis in clinical settings. Assessment varies between reviewers but in broad terms three primary components are typically considered: 1) exclusion of potentially treatable causes of dementia (e.g. mass-effect lesions, infection); 2) determining the extent of vascular disease; and 3) assessment of the degree and focality of cerebral atrophy. The first two of these components are not unique to the diagnosis of dementia and will be familiar to neuroradiologists. The third component, however, requires dementia expertise which is not readily available in every clinical setting. As a result, reports of cerebral atrophy can often be vague, without reference to many of the features with diagnostic value. Discussion of specific features associated with different neurodegenerative dementias is included in Chapter 3 but in broad terms, there are four key questions that can are commonly used to help define the cause of an individual's dementia:

Are the brain appearances appropriate for age? Typical ageing often results in a variety of brain changes without any discernible effect on cognitive function. These may include vascular changes, such as enlarged perivascular spaces and white matter hyperintensities [106], and/or cerebral atrophy often taking the form of enlarged ventricles, sulcal widening, or a degree medial temporal lobe atrophy [107, 108]. The assessment of brain imaging should, therefore, always take the person's age into consideration to determine if changes are likely to be pathological or just part of the typical ageing process.

Is there evidence of asymmetry? An asymmetric pattern of atrophy, either globally or locally (e.g. medial temporal lobe), is typically associated with FTLD pathology, and evidence of this should be reported to help guide clinical diagnosis [109, 110].

Is there an atrophy gradient? Evidence of an atrophy gradient, with greater atrophy in anterior rather than posterior brain regions is also an indication of an underlying FTLD pathology [109, 110]. Conversely, a posterior greater than anterior atrophy gradient is

more frequently associated with young-onset Alzheimer's disease [111]. Atrophy gradients may also be localised to within brain structures, such as the hippocampus, where greater atrophy anteriorly has also been reported in association with FTLD pathology [112].

Is there focal atrophy? Atrophy localised to within a particular lobe or brain region may also have positive predictive value for underlying pathology and evidence of disproportionate atrophy in particular brain regions should be included in the report [109, 110]. Specific features with positive predictive value for dementia diagnosis are discussed in Section 3.4.

2.2.2 Visual Rating Scales

Quantification of imaging features can enhance the diagnostic value of structural imaging but in most cases this requires specialist software and expertise so it is seldom used in clinical practice. By contrast, visual rating scales can be applied directly to clinically acquired images without the use of additional software, and with suitable training, can be used as an adjunct to standard clinical reads. Visual rating scales used to assess both subcortical signal changes and cerebral atrophy are discussed below. However, greater attention is paid to the cerebral atrophy scales as these are applied in practice to the work of this thesis.

Subcortical Signal Changes

The exact significance of white matter abnormalities demonstrated on T2-weighted or FLAIR MRI is not yet fully understood [113, 114] but by quantifying and localising the extent of these abnormalities in a standardised way, it is possible to look for correlations with demographic and clinical variables that may help to explain their contribution to cognitive decline. Visual rating scales have been designed with this purpose in mind, although since some include deep grey matter and brainstem regions, they are collectively referred to as subcortical signal change scales following STandards for ReportIng Vascular changes on nEuroimaging (STRIVE) guidance [115]. subcortical signal change scales are broadly divided into two categories. The first category includes scales designed (or adapted) to provide a general global assessment of vascular burden, and the second includes those designed to provide more fine grained assessment of specific

brain regions. The more general scales, such as the Fazekas [116] or Manolio [117] scales, focus on periventricular changes, appearing as "caps" and "bands" of high signal around the ventricles, and supratentorial white matter changes, ranging from punctate foci to large confluent areas. Of the more detailed scales, the Scheltens scale [118] and the Age Related White Matter Changes (ARWMC) scale [119] are widely used for research purposes. The Scheltens scale is the most complex, with a six-point scale applied in supratentorial, basal ganglia, and infratentorial regions, each of which are then subdivided into four smaller regions. The scale accounts for lesion size and frequency, and there is an additional three-point scale used to assess periventricular abnormalities [118]. The ARWMC scale is used to rate similar regions (minus the periventricular region) but using a simplified four-point scale, with the left and right hemispheres rated separately. The ARWMC scale was designed for use with both CT and MRI scans, however, inter-rater reliability was shown to be higher based on MRI scoring, and MRI was better able to detect smaller lesions [119]. The use of subcortical signal change scales is recommended in the latest STRIVE guidance [115]. A summary of the most commonly used rating scales discussed here is provided in Table 2.1.

Cerebral Atrophy

Visual rating scales have been developed specifically to rate several brain regions sensitive to atrophy in dementia, and can be used to provide quick, semiquantitative measures of the degree of atrophy in these regions [120]. They encourage structured image reporting and provide radiologists and non-radiology clinicians with a framework for interpreting imaging findings, making visual assessment more consistent and potentially more sensitive. Visual rating scales have been used extensively in research, and a review is provided in Table 2.2. A selection of these scales are discussed in detail below. More information on the design, methodology and validation of visual rating scales is included in Appendix B.

Global cortical atrophy: The global cortical atrophy scale by Pasquier *et al* was developed to evaluate atrophy on axial slices in 13 brain regions, including frontal, parieto-occipital, and temporal regions, and dilation of the ventricles [121]. Regions were assessed separately in each hemisphere and the final score was the sum of all scores in the thirteen regions. With such extensive coverage, the reliability of the original thirteen-

Region	Fazekas *et al* (1987) [116]	Scheltens *et al* (1993) [118]	Manolio *et al* (1994) [117]	Wahlund *et al* (2001) [119] (ARWMC)
Peri-ventricular	0=absence 1='caps' or pencil-thin lining 2=smooth 'halo' 3=Extension into WM	0=none 1=smooth halo ($\leq$1-5mm) 2=large confluent (5-10mm)	*Image-based references:* 0=none 1=barely detectable .. 8=extensive confluence 9=complete coverge	Not included
WMH	0=absence 1=punctate foci 2=beginning confluence 3=large confluent areas	0=none 1=<4mm; n$\leq$ 5 2=<4mm; n>5 3=4-10mm; n$\leq$ 5 4=4-10mm; n> 5 5=>10mm; n$\geq$ 1 6=confluent		0=none 1=focal lesions 2=beginning confluence 3=complete diffuse involvement
Basal ganglia	Not included		Not included	0=none 1=1 focal lesion ($\geq$5mm) 2=>1 focal lesion 3=confluent lesions
Infra-tentorial	Not included		Not included	0=none 1=focal lesions 2=beginning confluence 3=complete diffuse involvement
No. of citations	987	487	311	681

Table 2.1: Definitions for the most widely used subcortical signal change scales. The scales are designed to be performed on axial T2-weighted or FLAIR images. The number of times each scale has been cited is included as an indication of its impact. ARWMC=Age Related White Matter Changes, WM=white matter, WMH=white matter hyperintensities.

point scale was low and was further confounded by the inclusion of regions susceptible to partial volume effects. Subsequent simplification of the scale to a four point-scale based on a more global (not regional specific) impression of atrophy resulted in increased uptake among the scientific community (see Table 2.2), although the scale has been primarily used as a component part of larger diagnostic assessments [122]. Due to the large brain area assessed by global cortical atrophy scales, they are likely to be more severely confounded by age than other atrophy rating scales, although there is evidence to suggest this could be improved by using age-specific cut-offs [123].

Frontotemporal atrophy: Frontotemporal atrophy scales may be useful in the differential diagnosis of frontotemporal dementia syndromes, and the scales developed around these regions have been designed and validated specifically for this purpose. The most successful of these scales stemmed from a postmortem staging scheme used to rate atrophy in brains with FTLD pathology, which has been iteratively refined several times since [124]. The five-point scale was first devised by Davies *et al* [125] and rating was performed at the level of the anterior temporal lobe and the lateral geniculate nucleus. Kipps *et al* extended this scale further to include rating of the posterior temporal lobe, while also describing slice selection in greater detail, which resulted in improved reliability [126]. Finally, Ambikairajah *et al* refined the scale for use with patients of the MND-FTD continuum, which included rating of the orbito-frontal cortex, the anterior cingulate cortex and the motor cortex [127]. While references images were provided for each iteration of the scale, on close examination the spectrum of atrophy represented is not always uniformly distributed between scale increments. In some cases, such as the anterior temporal region, the scales may therefore benefit by being condensed to four points rather than five.

Medial temporal lobe atrophy: medial temporal lobe rating scales were first developed for use with CT imaging [128] but in the last twenty years they have been predominantly designed for use with MRI (see Table 2.2), as it became the imaging modality of choice in the diagnosis of dementia. The Scheltens scale [129] has had the biggest impact on the field, and has formed the basis of many subsequent scales [130–133], although none of them has had the impact of the original. The Scheltens scale focuses on three key features of medial temporal lobe atrophy, namely: the width of the

choroid fissure, the width of the temporal horn and the height of the hippocampus. The degree of atrophy in each of these regions is combined to produce a score reflecting overall medial temporal lobe atrophy. Both sides of the medial temporal lobe are assessed separately and in the case of asymmetry the highest score is reported. In order to assess sensitivity and specificity, the scale is dichotomised, with scores of 0–1 indicating the absence of Alzheimer's disease, and scores of 2–4 indicating the presence of Alzheimer's disease. Since it was introduced, the Scheltens scale has been used in over 100 studies (see Table 2.2). The reliability of the scale has been reported to be robust to the clinical experience of the rater [134] but increases as the rater gains more experience with the scale itself [135]. Better understanding of the pathological phenomenon measured by the scale has led to modification of the dichotomised scale to account for atrophy due to ageing [136, 137], which has also helped to improve performance. The Scheltens scale has been included in the research criteria for the diagnosis of Alzheimer's disease [138].

Posterior atrophy: There is currently only a single scale for the assessment of posterior atrophy, which was developed by Koedam [139] and Lehmann [111]. The scale focuses on the posterior cingulate sulcus, precuneus, parieto-occipital sulcus and the cortex of the parietal lobes. The left and right hemispheres are assessed separately and a separate score is given in each imaging plane (axial, sagittal, coronal). In the case of different scores in different planes, the highest score is taken. The scale is useful for the identification of young onset Alzheimer's disease, which often presents with posterior atrophy and relatively preserved medial temporal lobes. Combining both the medial temporal lobe scale and the posterior atrophy scale has been shown to have improved diagnostic accuracy for Alzheimer's disease [111].

2.3 Quantitative Assessment

A number of sophisticated analysis methods are available to quantify global and regional atrophy from MRI and they are used extensively in research, including clinical trials. Due to special hardware requirements, computational demands and dependency on specific acquisition techniques, relatively little progress has been made to date to integrate these into clinical work streams. This is likely to change in the near future

Scale	Brain region	Scale points	Imaging plane	MR contrast	Reliability		Citations	Applications	
					InterRater	IntraRater		Research	Trials
Pasquier [121]	Global cortical	4	Axial	T2-W	>0.6 (Cwκ)	>0.7 (Cwκ)	44	35	Y
O'Donovan [144]	Ventricles	4	Axial	T1-W	0.9 (ICC)	0.92 (ICC)	1	0	N
Davies/Kipps [125, 126]	Frontotemporal	5	Coronal	T1-W	0.62-0.71 (Cκ)	0.79-0.83 (Cκ)	90/60	9	N
Davies [145]		5	Coronal	T1-W	0.7 (Cwκ)	0.75 (Cwκ)	31	3	N
Ambikairajah [127]		5	Coronal	T1-W	0.9 (Uκ)	Not reported	0	0	N
Chow [146]		5	3-plane	T1-W	0.06-0.2 (Kw)	Not reported	0	0	N
De Leon [128, 147]	Medial temporal	4	Axial	T1-W	0.72 (Uκ)	Not reported	213	0	N
Scheltens [129]		5	Coronal	T1-W	0.72-0.84 (Cwκ)	0.83-0.94 (Cwκ)	350	100+	Y
Galton [130]		4	Coronal	T1-W	0.36-0.49 (Fκ)	0.8 (Cκ)	100	13	N
Urs/Duara [131, 132]		5	Coronal	T1-W	0.75-0.94 (Uκ)	0.84-0.93 (Uκ)	21/59	12	Y
Kaneko [148]		4	Coronal	STIR	0.68 (Uκ)	0.79 (Uκ)	0	0	N
Kim [133]		5	Axial	T1-W	0.64 (Uκ)	0.62/0.95 (Uκ)	1	1	N
Koedam [139]	Posterior	4	3-plane	T1-W FLAIR	0.65-0.84 (Cwκ)	0.93/0.95 (Cwκ)	19	5	N

Table 2.2: Atrophy visual rating scales developed or adapted for application to imaging in dementia. As an indication of the impact of each scale, the number of published studies that have applied them is provided, as well an indication of their inclusion in clinical trials. Highest reported reliability values are listed - the citation is listed if not taken from the original paper. Search carried out April 2014. Abbreviations: Cwκ=Cohen's weighted κ, Cκ=Cohen's κ, Fκ=Fleiss' κ, ICC=intraclass correlation coefficient, KW=Kendall's W, Uκ=Unspecified κ, STIR=Short T1 Inversion Recovery, T1-W=T1-weighted, N=no, Y=yes

and there is already work under way to integrate some of these techniques directly with hospital picture archiving and communications systems (PACS). In this section some of the key steps for quantitative image assessment are discussed, in part to help inform the analysis decisions taken in relation to the work of this thesis.

2.3.1 Correcting Intensity Non-Uniformity

intensity non-uniformity (INU), or 'bias', is a common artefact in MR images that results in a smooth intensity variation within an image. While this has limited effect on visual assessment it can impact significantly on image registration and segmentation algorithms, which often assume that a single tissue type is represented by similar voxel intensity [140]. The artefact can be caused by multiple factors related to scanner hardware, pulse sequence or the object being imaged (reviewed in [141]).

Both prospective and retrospective correction techniques are available. Prospective techniques focus on correcting factors related to the acquisition process by acquiring images with a uniform phantom, multiple coils, or special pulse sequences [142]. Retrospective correction is based directly on the intensity information extracted from the acquired image, which in some cases is integrated with a priori spatial and/or intensity probability information about the imaged anatomy [142]. Unlike the prospective techniques, these techniques are also able to correct INU due to the object being imaged (i.e. the head) and are more frequently used in neuroimaging studies [143].

2.3.2 Image Registration

Image registration is a crucial first step in the quantitative assessment of medical images, where the extraction of valuable information is based on more than one image. Examples include the fusion of multi-modality images to enhance diagnostic value, assessing change between multiple timepoints for disease monitoring, determining differences at a group level to identify distinguishing features, and the construction and utilisation of atlas based techniques for volume extraction [149]. Mathematically, image registration is an iterative process that involves the alignment of two or more images based on a geometric transformation, a subsequent measure of how well the images are aligned, and a means of optimising the process to achieve a good solution.

Image transformations can be categorised as affine or non-rigid based on the extent of image distortion used to align the image with a target. Affine registrations in 3D space have a maximum of twelve degrees-of-freedom (DF) to allow translations, rotations, scaling and skews, whilst preserving all parallel lines in the image. A special case of affine registration is rigid registration, which includes only translations and rotations (six DF), and preserves all distances in the image [149]. Affine registrations are particularly useful for aligning bone and in neuroimaging are often used to provide approximate alignment. To account adequately for the degree of variation in brain tissue, however, more complicated deformations are often required. non-rigid registration transformations can provide polynomial mapping between the coordinates in the original image and the target image, and many more degrees of freedom to allow more flexible distortion of the image. non-rigid registration algorithms may be based on the alignment of image landmarks or the matching of image intensity [149].

The extent of image deformation should be constrained to ensure a biologically plausible solution. This is typically achieved by implementing a cost function which combines a post-transformation similarity measure with a regularisation term, modelled on a physical process such as fluid flow, that penalises unlikely transformations such as folding or high local stretching or bending [149]. Additional constraints include a trade-off between computation time and accuracy, which is influenced by factors such as the DF and the number of iterations. Image pre-processing, such as bias correction (discussed above) and constraining the algorithm to a masked region, can also improve the quality of the registration.

2.3.3 Image Segmentation

Image segmentation is used to simplify the representation of an image into anatomically relevant classes to allow meaningful analysis of brain tissue. Segmentation methods assign a label to each voxel such that voxels with the same label are assumed to share anatomical or physiological characteristics. A selection of the methods used for image segmentation are reviewed in this section.

Manual Delineation

Manual delineation of brain structures is considered the gold standard in image segmentation as a human operator is better able to adapt to variations in image quality and/or anatomy that can potentially confound automated algorithms based on predefined rules. Manual operators, however, are not without error and a degree of both inter- and intra-operator variation does exist, introducing uncertainty into the resulting volume estimates [150]. Segmentation protocols for various different brain regions have been developed to help standardise the process and minimise measurement error, however, with multiple published protocols used for different studies, considerable variation still exists [150, 151]. To address this issue for hippocampal segmentation, which is often outcome measure in clinical trials, the European Alzheimers disease consortium (EADC) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) are currently developing a harmonised segmentation protocol based on the most highly cited protocols in the literature [152].

Despite advantages in segmentation accuracy, manual delineation is time-consuming, labour-intensive and expensive, and unlikely to be adapted into routine clinical use. It does, however, continue to be used to validate automated segmentation techniques [153].

Automated Techniques

Pre-Processing Steps

As mentioned briefly above, there are several undesirable yet often unavoidable features of an image that can confound automated segmentation algorithms, such as artefacts, scan orientation, and anatomical variation (exacerbated by disease pathology). Image pre-processing is, therefore, essential to ensure a good starting point for the algorithm. Key image pre-processing steps includes INU correction (as discussed above in Section 2.3.1), image registration (as discussed above in Section 2.3.2), and removal of non-brain tissue.

Image registration prior to segmentation is typically to a standard template, such as the International Consortium for Brain Mapping 152 (ICBMI152) template [154], which is the average of 152 T1-weighted images acquired in healthy adult brains approximately

transformed to Talairach space [155] (for more details on its evolution see [156, 157]). This aligns images in a standard coordinate space, providing approximate voxel-wise correspondence, to allow subsequent mathematical operations to be applied between images and comparisons to be drawn between results.

Non-brain tissue such as the eyes, dura and skull have signal intensities similar to brain tissue, therefore, their extraction reduces the segmentation to a simplified anatomical region [158]. Techniques to remove non-brain tissue produce either a new image containing only brain voxels or a binary mask (1 for brain voxels, 0 for non-brain tissues), which can be incorporated into the segmentation process.

Segmentation Methods

Various segmentation methods have been implemented to answer a diverse range of clinical questions with medical imaging. These methods can be classified into four categories discussed below.

Intensity-based methods classify each voxel based on its signal intensity. These methods include thresholding, which uses the intensity histogram to determine the optimal threshold to best separate specific tissue classes; or region growing, which starts from a seed point and extends to include all voxels with similar intensities. Intensity-based methods are generally suitable for the classification of grey matter, white matter and CSF, but unsuitable for segmenting smaller brain structures that have substantial overlap in intensity. These techniques are particularly sensitive to noise in the image [153].

Atlas-based methods use (often manually) segmented brain images as a template for segmenting new images. In this context an individual brain atlas typically consists of an intensity image (e.g. T1-weighted image) and a corresponding segmented or labelled image. Atlases are based on a single subject thought to represent average anatomy and signal intensity, or on a population average thought to be more representative of natural anatomical variability. The image to be segmented is typically aligned to the atlas in a two stage process. Firstly an affine registration is used to achieve global alignment at a low computational cost. Secondly a non-rigid registration is used to achieve local alignment of specific anatomy at a much higher computational cost. Accurate registra-

tion is vital to allow the spatial information to be transferred directly from the atlas to the image, allowing multiple brain structures to be segmented at once, which is a major advantage of this technique. Anatomical variation due to disease pathology introduce additional challenges, however, and ideally the atlas used should be representative of the population under study. Currently most available atlases are based on a healthy control population [159].

Surface-based methods attempt to fit a parameterised surface (or mesh) to a region of interest whilst maintaining certain shape characteristics. These techniques combine rules from geometry, to represent shape, and physics, to define how the shape deforms. These methods are commonly used to determine cortical thickness, where an initial mesh is generated from whole brain segmentation and evolved to fit the cortex, with constraints applied based on prior anatomical knowledge [160].

Hybrid methods combine elements of the other segmentation methods discussed, to achieve an optimal solution. There is no ideal segmentation solution for all tasks, or even for any one task, however, the benefit of a hybrid approach is that it can use a combination of methods to compensate for the disadvantages of any one technique. Most state-of-the-art techniques now developed for neuroimaging are examples of hybrid techniques [153].

2.3.4 Voxel Based Morphometry

voxel-based morphometry (VBM) is a tool for studying differences in composition of brain tissue between groups, and can also be used to investigate correlations between subject characteristics and neuroanatomical patterns. The technique provides a registration and segmentation pipeline to pre-process brain imaging data in such a way as to sensitise it to a clinical signal of interest. It then allows the user to specify a linear combination of features to explain this signal, modelled (using the general linear model) at the voxel level. These models can then be interrogated statistically to look for group differences or clinical correlations in the image signal. Although there are other VBM impementations available (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLVBM), this technique will be described here in terms of the widely used statistical parametric mapping (SPM) framework [161].

Unified Segmentation

Unified Segmentation, the first step in the VBM processing pipeline, is a probabilistic framework, developed by Ashburner and Friston, that allows image registration, segmentation and bias correction to be combined within the same model [162]. By integrating these three steps in a single model, the final solution is optimised rather than the sequential steps. This technique is used to create roughly aligned grey and white matter segmentations for each image that will be refined in subsequent steps of the processing pipeline. The available (user-specified) options to tune the model are described below.

Bias correction: In SPM, INU (described in Section 2.3.1) is modelled as a linear combination of smooth basis functions. The bias correction regularisation term in SPM allows prior information about the degree of INU in the images to be introduced into the model. Increasing the regularisation term indicates that the image contains less INU and reduces its influence over the model. An option is also provided to set the full width at half maximum of the Gaussian smoothing kernel. A large full width at half maximum should be chosen if there is less INU to prevent loosing the natural intensity variation between the different tissue types.

Tissue classification is performed by fitting a six component Gaussian mixture model (representing grey matter, white matter, CSF, bone, soft tissue, air) to each image intensity histogram [162]. Based on a combination of the location (given by the mean), the spread (given by the variance) and a weighting factor for each Gaussian, a reasonably accurate model of the data can be obtained. By associating each of the Gaussians with a different tissue type it is possible to calculate the probability of a particular tissue type given the image intensity at that voxel. SPM provides the option to extend this model to fit multiple Gaussians per tissue class, relaxing the assumption that a single tissue type has a Gaussian intensity distribution, which can help to improve the accuracy of the model.

Starting Estimates: SPM provides an option to specify a tissue probability map, or brain atlas (as described in Section 2.3.3 - Atlas-based methods). All images are roughly aligned to the tissue probability map coordinate system using a twelve DF

affine transformation to provide an initial starting point for the iterative process of fitting the model. Further regularisation of deformations can be specified to improve alignment, however, the default sttings are typically used if Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) is included in the VBM processing pipeline (described below). It is also possibe to set the distance between the points sampled during model fitting. Smaller values use more of the data and produce more accurate results but are computationally more expensive and take longer to run (www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf).

Clean up process: By setting the "cleanup" parameter, misclassified voxels are crudely accounted for using dilations and erosions of the segmentations. The level of cleanup specified constrains the extent of the dilations and erosions. Additional removal of misclassified voxels is achieved through a few iterations of a simple Markof random field (MRF) model (http://dbm.neuro.uni-jena.de/vbm/markov-random-fields/). The technique compares the classification of the central voxel in a neighbourhood of voxels (in this case a 3x3x3 neighbourhood) and, based on the classification of its neighbours, combined with the prior probability of the class, determines the likelihood of the central voxel being correctly classified. The prior probability is determined by the strength of the MRF, a weighting of zero indicates no MRF, while higher weightings can be used for noisy data. This additional clean up procedure is designed to deal with isolated voxels or holes in a cluster of connected voxels that are likely to have been misclassified.

DARTEL

DARTEL is used to estimate the deformations that best align the roughly segmented grey and white matter segmentations generated by the unified segmentation framework [163]. By incorporating many more parameters into the representation of brain shape (3 DF per voxel), DARTEL increases the flexibility of the registration process to achieve more accurate inter-subject alignment.

Starting with the roughly aligned segmentations generated by the unified segmentation process, group average templates are generated by simultaneously aligning grey matter with grey metter and white matter with white matter. As DARTEL enforces a

one-to-one mapping between the forward and backward deformations, this initial group template is iteratively refined by computing the deformations from the template to each of the individual images and averaging the inverse transformations applied to each image [163]. With each iteration an increasingly crisp template image is generated. Flow field images representing the direction of the deformation of each image to the final template are also generated.

Final Segmentations

The final group template generated by DARTEL is affinely registered to the specified tissue probability map. The parameters of this registration are used, along with the computed flow fields from the DARTEL transformations, to warp the native space grey and white matter segmentations generated by the unified segmentation model [164]. Signal intensity is modulated based on the degree of deformation in each brain region, with areas of expansion reduced in intensity to preserve the total amount of signal in the image [164]. The images are then convolved with a Gaussian smoothing kernel, which sensitises the analysis to a particular spatial scale of effect determined by the full width at half maximum of the kernel. The value of the full width at half maximum depends on the accuracy of the inter-subject registration to the tissue probability map space, with more smoothing required for less accurate registrations. The optimal value is likely to vary between brain regions with highly variable cortical regions requiring more smoothing and less variable subcortical regions requiring less smoothing [164].

Masking

Voxel-wise statistical analysis on the final segmentations should include only voxels contained within a specified mask to help reduce the number of comparisons, and to reduce the number of false positive results. While smaller masks can increase sensitivity and help clarify interpretation, if the mask is overly restrictive potentially interesting regions of interest will be excluded from the statistical analysis resulting in an increase in false negative results.

Ridgway *et al* developed masking strategies to help generate an optimal mask for use in studies of neurodegenerative disease; these strategies are now available as an SPM toolbox (http://www0.cs.ucl.ac.uk/staff/g.ridgway/masking/) [165]. Prior to this work,

masks were typically generated in SPM by excluding voxels in which any of the images had intensity values below a certain threshold, with the threshold specified as a constant value or set as a fraction of each images global value. The intersection of the individual binary masks was then found to create the group mask. However, patients with neurodegenerative dementia often have overlapping patterns of atrophy, particularly in the medial temporal lobe, therefore, by adopting this strategy clinically interesting regions may be excluded from the analysis. The methods developed by Ridgway *et al* addresses this issue in two ways. In one method this is achieved by relaxing the criteria that all subjects are required to have voxel intensity above a certain threshold. Instead voxels are included in the mask if there is consensus among some percentage of subjects that their intensity is above the threshold. In the other method, a threshold is applied to the mean intensity mask created from all subject's segmentations. The mean intensity mask is highly likely to contain all relevant regions of interest. A good threshold applied to this average image should remain highly correlated with the unthresholded original. By using a function to maximise the Pearson correlation coefficient over pairs of voxels between the unthresholded mask and a thresholded mask it is possible to maximise the threshold objectively, independently of the operator. This second strategy is therefore recommended in the first instance [165].

Statistical Analysis

Using the general linear model, SPM provides a framework for specifying a mixture of variables to define the processed image data [166]. By using classical statistics to interrogate the derived model it is possible to infer which variables contribute significantly to any within- or between-subject variance.

Defining the imaging data in terms of the general linear model decomposes the data into effects and error based on the following equation:

$$y = X\beta + \varepsilon$$

Where y is the signal intensity in the image data, X is the design matrix containing all available knowledge about the experimentally controlled factors and the potential confounds, β is the weighted estimate of the contribution of the design matrix parameters

(regression coefficients), and ε is the residual signal that is not explained by the design matrix. Ideally the estimated parameters should explain the signal as much as possible, and in doing so minimise the unexplained variability represented by the error term. The errors are typically assumed to be normally distributed with independent and identical noise across the images. The parameters are estimated by minimising the squared error term using ordinary least squares estimation.

EXAMPLE

Consider a VBM experiment to compare grey matter density between AD patients and controls. The general linear model equation above could be rewritten as:

$$y = \begin{bmatrix} x_{1AD} & x_{1Controls} & x_{1Age} & x_{1Gender} & x_{1TIV} \\ x_{2AD} & x_{2Controls} & x_{2Age} & x_{2Gender} & x_{2TIV} \\ x_{3AD} & x_{3Controls} & x_{3Age} & x_{3Gender} & x_{3TIV} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{nAD} & x_{nControls} & x_{nAge} & x_{nGender} & x_{nTIV} \end{bmatrix} \begin{bmatrix} \beta_{AD} \\ \beta_{Controls} \\ \beta_{Age} \\ \beta_{Gender} \\ \beta_{TIV} \end{bmatrix} + \varepsilon$$

Where X_*iAD* and X_*iControls* are binary indicator variables denoting group for each subject (1-n) and signal intensity is corrected for the confounding factors of age, gender and total intracranial volume (TIV) (discussed in Section 4.4.3) by including them as covariates in the model [167]. A contrast vector, c, can then be defined by selecting a linear combination of regression coefficients to investigate the effect of interest, e.g. are there differences in grey matter density between the Alzheimer's disease group and the control group.

$$c^T\hat{\beta} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{AD} \\ \beta_{Controls} \\ \beta_{Age} \\ \beta_{Gender} \\ \beta_{TIV} \end{bmatrix} \quad Or \quad \beta_{Controls} - \beta_{AD}$$

The null hypothesis, H_0, is therefore specified as:

$$H_0 : c^T\hat{\beta} = 0$$

H_0 can then be interrogated using univariate statistics such as the simple t-test, defined as the contrast of estimated parameters divided by the square root of the variance estimate.

$$T = \frac{c^T \hat{\beta}}{\sqrt{var(c^T \hat{\beta})}}$$

VBM applies such tests at the voxel level and a statistics image is generated over the entire brain volume. Analysis using a standard structural T1-weighted volumetric scan will, therefore, result in over 1 million tests at the group level, with every test increasing the likelihood that the groups will differ in some regions by chance alone. Multiple testing correction methods are available to control the chance of false positive results, essentially setting a threshold below which an effect is considered noise. family-wise error rate (FWE) and false discovery rate are commonly used in neuroimaging studies for this purpose [168, 169]. FWE is the probability of any false positives occurring in any individual tests, while false discovery rate corrects for a proportion of false positives among the rejected null hypotheses. FWE is, therefore, the more stringent method of correcting for multiple tests. These methods are implemented in the context of random field theory due to the spatial correlation between neighbouring voxels (particularly after smoothing) [166]. This violates the assumption of independent statistical tests assumed by the more standard Bonferroni correction, and would result in an overly conservative threshold if this technique was applied [170].

2.3.5 Machine Learning

Machine learning algorithms quantify relationships within existing data and use these identified patterns to make predictions based on new data [171]. Increasingly, these algorithms are used to interrogate a vast array of medical data, including imaging data, to look for patterns previously undetected by more conventional statistical analysis or the human eye. The basic concepts of machine learning for practical implementation are discussed below.

Learning Algorithms

Learning algorithms can be separated into two categories, supervised or unsupervised.

- Supervised learning algorithms are used when the existing data is already labelled and the algorithm is required to classify the new data in the same way. This type of learning problem can be then categorised as a classification problem, if the data can be separated into discrete classes, or a regression problem, if the desired output is a continuous variable. Popular supervised learning algorithms include support vector machine (SVM), k nearest neighbour classifiers and random forests.

- Unsupervised learning algorithms are used to discover clusters within the data when no target value is available. Popular unsupervised learning algorithms include k-means, Gaussian mixture model, and heirarchical clustering.

The choice of learning algorithm depends not only on the data available and the question to be answered but also on the number of samples available within dataset. Machine learning algorithms are best suited to large datasets often with several thousand samples, and in some cases unsuitable for smaller datasets below this order of magnitude.

Training and Testing

In order for an algorithm to learn patterns in the data it needs to be trained using a data sample representative of the complete study population, and tested using a similarly representative sample. Taking a classification problem as an example, the training and testing dataset must contain each class defined in the problem, and both should contain the same proportion of classes. A dataset with unbalanced classes will bias the algorithm towards the larger class, however, if this is unavoidable a weighting factor is applied to increase the cost of misclassifying the smaller class and reduce this bias. Both training and testing sets must be mutually exclusive to prevent optimising the algorithm for the study population, which will fail to generalise to the wider population, a process referred to as over-fitting.

The Classification Problem

Machine learning datasets are typically multivariate, with each sample in the dataset (i..N) defined by an array, x, of M features. In a classification problem, the solution,

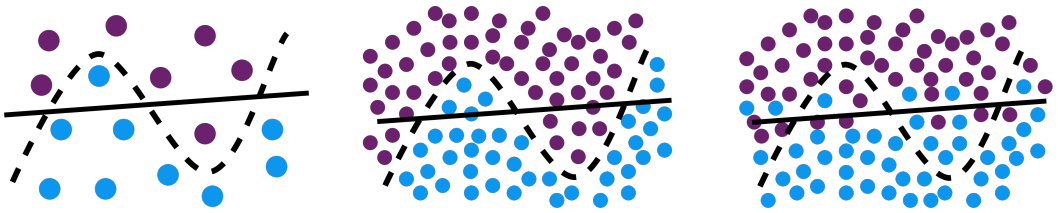


Figure 2.1: Given only a small sample (left image) either, the linear (solid) or non-linear (dashed) boundary might be more appropriate. A larger sample (middle and right image) better reflects the true distribution. If the middle image represents the true distribution, the linear boundary under-fits the data. If the image on the right represents the true distribution, the non-linear boundary over-fits the data. Diagram recreated from Lemm *et al* [172]

y, can be specified as a function of x, i.e. $y = f(x)$, and each data sample can be represented as a point in n-dimensional space. Taking a binary classification problem as an example, the goal of a classifier is to construct a boundary in n-dimensional space that classifies the samples into two classes. This boundary can either be linear or non-linear, as shown in Figure 2.1. In some cases a non-linear boundary is more appropriate, however, as shown in Figure 2.1, a non-linear boundary may over-fit the data given only a small sample size. The popular linear SVM algorithm, often used with smaller sample sizes, is described here and used for the work of this thesis.

Support Vector Machine

An SVM can be implemented as a linear classifier to construct a separating hyperplane, T (n-1 dimensions), that classifies all data into two classes. The goal is to position T such that the distance (margin) to the closest data points on either side (the support vectors) is maximised. Focusing the solution on the support vectors reduces the influence of outlier data points and produces better classification accuracy.

Taking the simple case, based on two classes, where x has only two features as an example, $x = (x_1, x_2)$, and T is a two dimensional line, the discriminant function can be described as

$$f(x) = w^T x + b$$

Where f(x), has a value of zero at T, a value less than zero for class 1, and a value greater

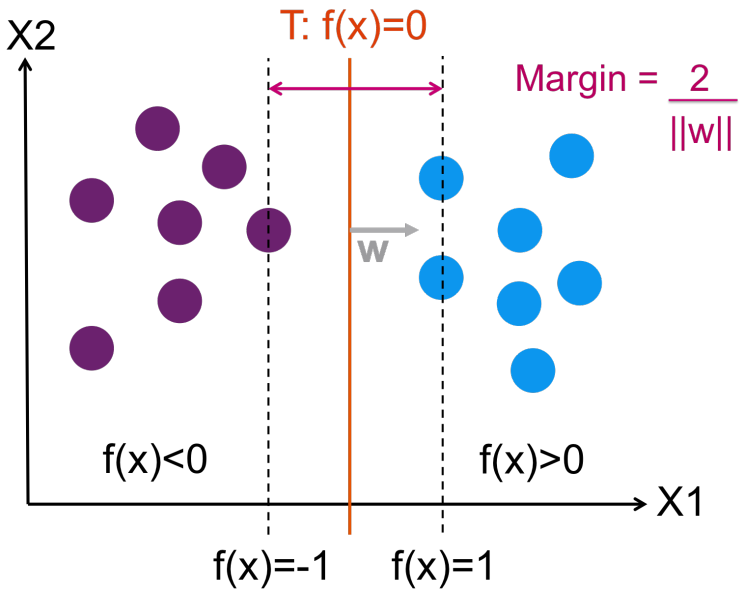


Figure 2.2: The discriminant function, f(x), has a value of zero on the separating hyperplane, T. For data points defined as class 1 (purple dots) f(x)<0, for class 2 (blue dots) f(x)>0. The goal of the SVM is to define T to maximise the margin between the support vectors in class 1 and class 2. W is the weight vector perpendicular to T.

than zero for class 2 (see Figure 2.2). W, the weight vector perpendicular to T and b is the bias term, which dictates whether T goes through the origin of the coordinate system (b=0 represents an unbiased hyperplane through the origin).

By normalising w such that $f(x_{class1})$=-1 and $f(x_{class1})$=1 the margin is given by

$$margin = \frac{1}{||w||} + \frac{1}{||w||} = \frac{2}{||w||}$$

Minimising the weight vector will, therefore, maximise the margin and the separability of the data. Misclassifications are inevitable in most cases and the optimal solution is a trade off between maximising the margin and minimising the number of misclassifications.

The extent of this trade off can be mediated using the SVM regularisation parameter, c. If a wider margin is the preferred outcome (with a risk of under-fitting to the training data), a smaller c value should be selected, and vice-versa if minimal misclassification is the preferred outcome (with a risk of over-fitting to the training data). If confident that the training data is truly representative of the general distribution, higher c values

may produce better classification accuracy, however, in most cases the optimal choice of c value is best determined empirically using a technique known as grid-search. Using this technique a range of c-values are tested for every iteration of training process and the value that consistently optimises the algorithm for the desired outcome is selected.

As previously mentioned, avoiding over-fitting the algorithm to the training data is a challenge in machine learning. To reduce this risk the training process can be considered as a sub-problem of the larger classification problem, where the training data is further divided into training and testing sets. This technique is known as cross-validation. A special case of cross validation, known as leave-one-out, trains on N-1 samples and tests with one sample a total of N times.

2.4 Conclusion

Many methods are available to extract clinically valuable and complementary information from structural MRI. Simple visual assessment techniques are quick and easy to apply and are well suited for implementation into clinical practice. More advanced techniques allow for quantitative measures to be extracted which may help with more fine-grained analysis. However, these advanced techniques are often complex and require specialist software, hardware and expertise. As computing capabilities continue to advance, these more advanced techniques, that have been used extensively in dementia research, will eventually be integrated into clinical workstreams. In this thesis, several of these techniques, from simple to complex, are investigated for their ability to characterise and distinguish between imaging data acquired in pathologically proven dementias, data which is rarely available in dementia research.

Chapter 3

Evidence for the Use of Structural MRI in the Diagnosis of Dementia

3.1 Introduction

As discussed previously, structural MRI serves multiple purposes in the clinical assessment of cognitive impairment. Firstly, it can be used to exclude potentially treatable causes of dementia (e.g. mass-effect lesions, infection), secondly to determine the extent of vascular disease, and thirdly to identify characteristic patterns of cerebral atrophy. These three broad categories are included together, or at least in part, in consensus diagnostic guidelines for all neurodegenerative dementias, and make up a general framework for image assessment (see Figure 3.1). Although the subject of this thesis is the diagnostic value of cerebral atrophy, the imaging features at each stage of the assessment process are discussed in this chapter due to the important implications of each in the diagnosis of dementia and the role of structural MRI in their identification.

3.2 Exclusion of Mass Effect Lesions

Brain lesions of a certain size or within a certain location may be sufficient to cause cognitive impairment. Such lesions are potentially suitable for surgical intervention and their identification should be the starting point when reviewing structural imaging. These include tumour (eg, meningioma, glioma), subdural haematoma, arteriovenous

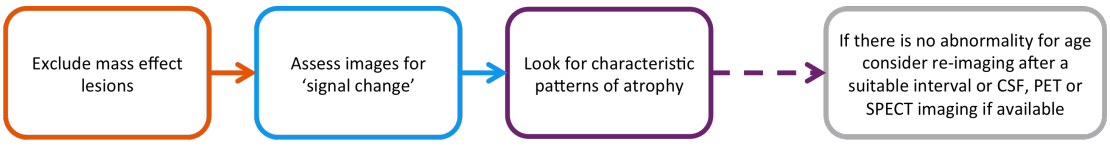


Figure 3.1: A general framework for image assessment in cognitive impairment, previously proposed by Barkhof *et al* and Harper *et al* [69, 173]

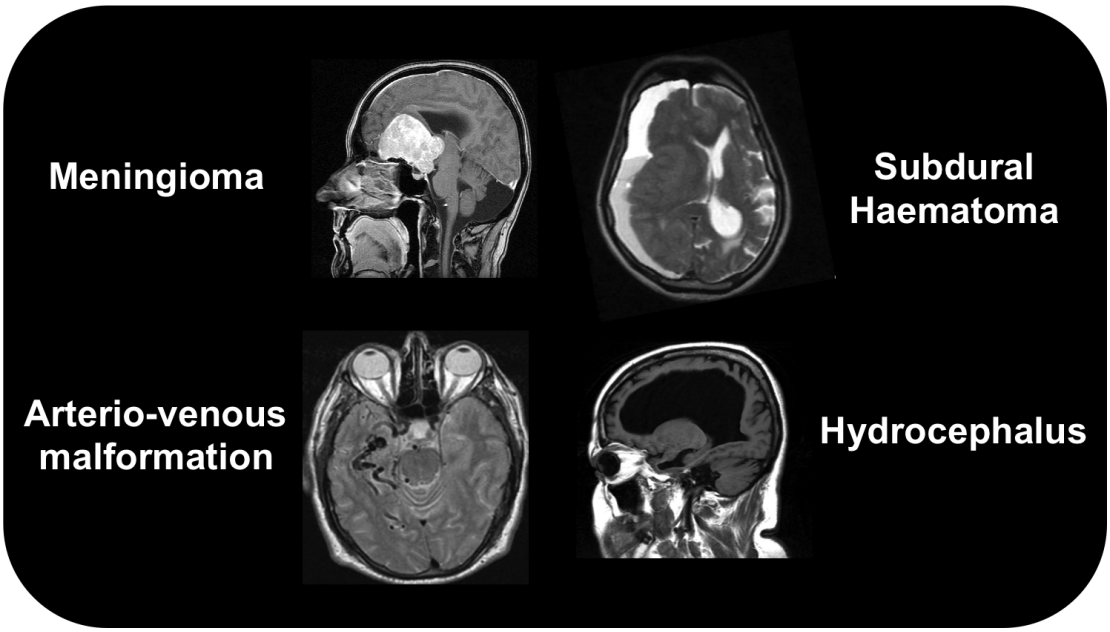


Figure 3.2: Lesions of a certain size, in a certain location may be sufficient to cause cognitive impairment. The subdural haematoma image and the hydrocephalus image was provided by Radiopaedia.org

malformation and hydrocephalus (see Figure 3.2). Radiology expertise is essential at this stage, particularly for small or isodense lesions that may be missed by an untrained eye.

3.3 Assessing Signal Change

Signal intensity within a single tissue type should be reasonably uniform on MRI and the presence of regions of hyperintensity or hypointensity within a single tissue type typically reflects pathology. In the context of dementia, MR signal change is frequently associated with vascular disease; however, it may also indicate inflammatory, metabolic, toxic or infective processes which could be contributing to cognitive deficits.

Vascular disease and Alzheimer's disease share many risk factors, including an increasing prevalence with age, and in many patients these pathologies coexist [174, 175]. The exact interaction between neurodegenerative and vascular pathology is debated but critically for the patient, it is important that vascular risk factors are managed and treated. Differentiating the relative contribution of vascular from neurodegenerative pathology as the cause of a patient's cognitive problems is a common clinical problem where MRI

can play a central role.

vascular cognitive impairment or vascular dementia is most frequently associated with atherosclerosis of the cerebral arteries, cerebral small vessel disease (SVD) and CAA [176]. These disorders manifest as cerebral infarcts, white matter lesions and haemorrhages, which can be visualised on structural MRI and will be discussed below in terms of small or large vessel involvement.

3.3.1 Large Vessel Disease

Atherosclerosis leads to the accumulation of blood derived lipids and proteins, particularly cholesterol, within the vessel wall [176]. This causes plaque formation and calcification, resulting in a reduction in blood flow to brain tissue. Complete or prolonged ischaemia causes focal regions of tissue necrosis, i.e. cerebral infarctions, which typically evolve into fluid filled cavities as interstitial fluid accumulates in and around the infarction. The altered tissue composition means that these lesions are readily visualised with MRI.

A diagnosis of vascular dementia due to large vessel disease is based on both the topology and severity of infarcts evident on T2-weighted or FLAIR images as (operationally) defined by the NINDS-AIREN (National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences) criteria for vascular dementia [177, 178]. Infarcts must be located in vascular territories around the anterior (bilaterally), middle or posterior cerebral artery, or in associated watershed regions (see Figure 3.3). In terms of severity, infarcts must be present in the dominant hemisphere or bilaterally, at least meeting the topography criteria in the non-dominant hemisphere.

There is evidence to suggest that the association between large vessel infarcts and cognitive impairment is weaker than the same association with small vessel infarcts. However, any previous infarction carries a greater risk of developing dementia [180, 181]. The presence of cerebral atherosclerosis, independent of infarction, and therefore potentially reversible, has also been shown to increase the risk of cognitive impairment [182, 183].

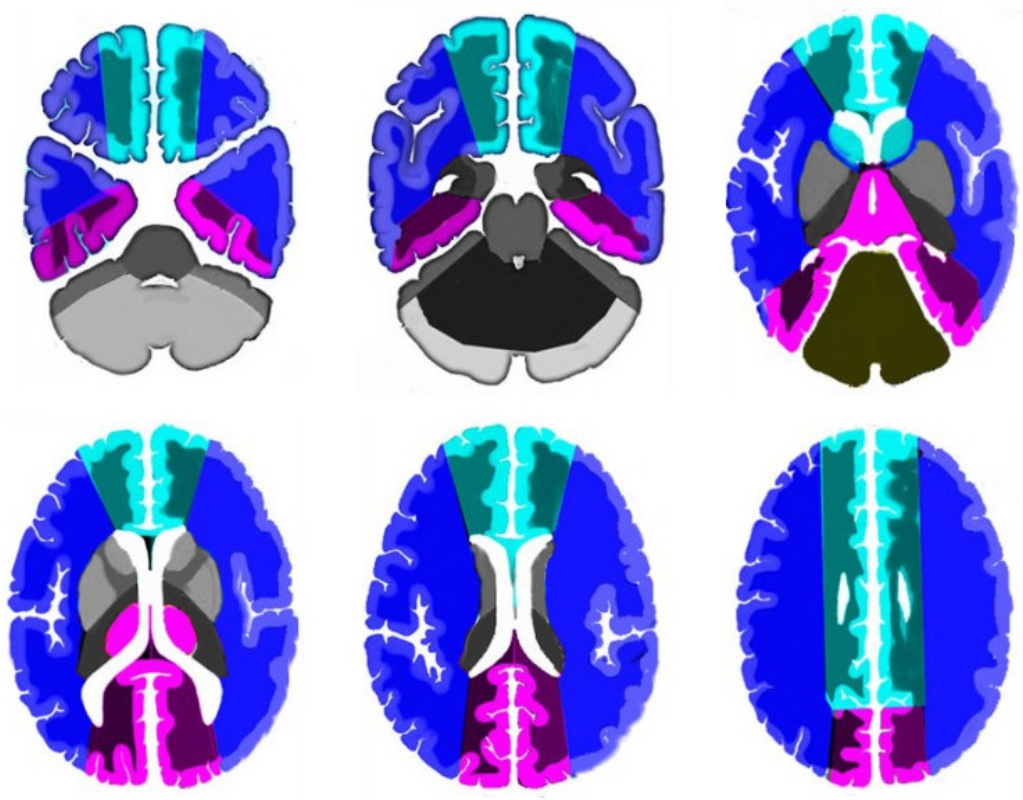


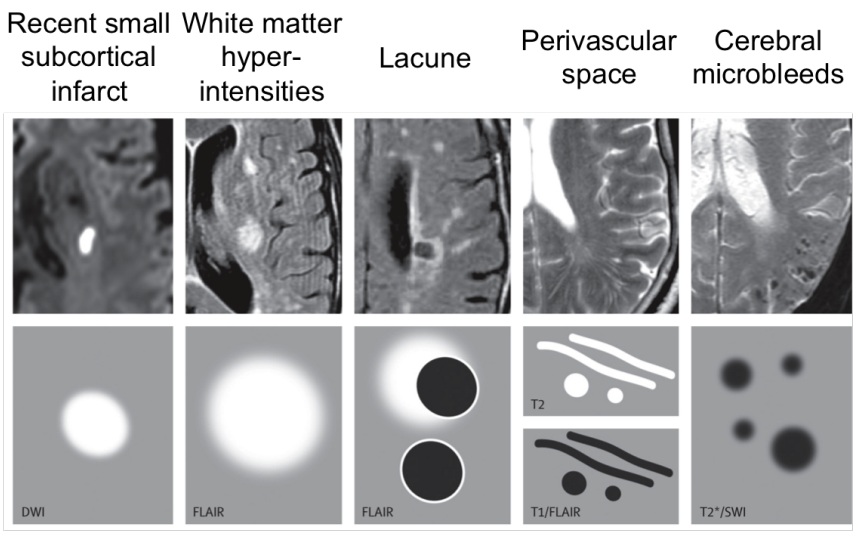
Figure 3.3: Vascular territories of the cerebral arteries. Anterior cerebral artery territory is shown in turquoise, middle cerebral artery in blue and posterior cerebral artery in magenta. The diagram was adapted from the Radiology Assistant website (http://www.radiologyassistant.nl) and originated in Savoiardo *et al* [179]

3.3.2 Small Vessel Disease

Small vessel disease describes a number of discrete pathological entities that affect the small vessels of the brain. These are described below in terms of their appearance on MRI using the suggested categories from the recent STRIVE criteria [115] (see Figure 3.4).

Lacunes of Presumed Vascular Origin

Arteriosclerosis and lipohyalinosis are common causes of SVD that lead to thickening of the vessel walls but without the additional calcification that occurs in atherosclerosis of the larger vessels [184]. Cerebral infarction or haemorrhage often occur as a result, however, the size and location differ from the larger vessel equivalents as a result of their vascular origin. Pathological staging of Arteriosclerosis/lipohyalinosis in relation



Typical diameter	≤20mm	Variable	3-15mm	≤2mm	~≤10mm
Best sequence	DWI	T2/FLAIR	T2/FLAIR	T2/FLAIR	T2*/SWI
T1	↓	↔ ↓	↓	↓	↔
T2	↑	↑	↑	↑	↔
FLAIR	↑	↑	↓	↓	↔
T2*	↔	↑	↔ ↓*	↔	↓

Figure 3.4: MRI findings related to small vessel disease - recommendations from STRIVE [115]. *If the lacune is the result of a haemorrhage.
This table is adapted from [115] with permission from Elsevier.

to cognitive deficits has been proposed to start in arteries of the basal ganglia, expanding into peripheral white matter and leptomeningeal arteries, then into thalamic and cerebral white matter vessels, and finally into brain stem arteries [22].

Radiologically, these smaller infarcts are referred to as lacunes and defined as small, round or ovoid, fluid-filled cavities in the region of a perforating arteriole. They are typically between 3-15mm in diameter with CSF-like intensity on T1-weighted, T2-weighted and FLAIR images, although they are best visualised on the latter two sequences. On FLAIR images lacunes often have a hyperintense rim, which helps to distinguish them from prominent perivascular spaces (see below). Bilateral thalamic infarcts however, which are sufficient to cause cognitive impairment [178], are better identified on T2-weighted than FLAIR images [185]. Ideally both sequences should be

acquired for thorough investigation of small vessel disease.

There is evidence to suggest that not all infarcts evolve in the same way [186, 187] and in some cases conventional MRI sequences are not sensitive enough to detect recent infarcts. Diffusion weighted imaging is recommended for the identification of subcortical infarcts occurring within weeks, which appear as hyperintense lesions, potentially larger than the 15mm upper bound previously defined for lacunes. Improved sensitivity means there is also no lower size limit given for infarcts detected on diffusion weighted imaging [115].

Lacunes are associated with an increased risk of stroke, gait impairment, and dementia, with increasing prevalence in relation to advancing age, hypertension, diabetes and hypercholesterolemia [180, 188–190].

White Matter Hyperintensities

The pathogenesis of white matter lesions, visible as hyperintensities on MRI, is complex and not yet fully understood [191]. In the context of dementia they are most frequently associated with small vessel disease and characterised pathologically by demylination, axon loss, astrogliosis and microglia activations [176].

On T2-weighted and FLAIR sequences the severity of white matter hyperintensities can be visually quantified by application of an established rating scale as discussed in Section 2.2.2. These scales typically classify lesions by location as periventricular, deep white matter hyperintensities or deep grey matter (if the latter case is included the collective term is subcortical hyperintensities [115]), although attempts to clarify the clinical consequences of each have been contradictory [192, 193]. Confluence of hyperintensities in at least two brain regions, and the beginning of confluence of hyperintensities in a further two regions, is considered to represent the involvement of at least a quarter of the total white matter and is sufficient for a diagnosis of vascular dementia due to SVD [178]. However, even in cases of extensive white matter hyperintensities, the existence of mixed pathology should be considered, although it may be difficult to confirm or refute. Less often, multifocal/confluent regions of hyperintensity in a patient with suspected dementia may result from a number of other conditions, including infections, inflammatory demyelinating diseases, leukodystrophies or

leukoencephalopathies [69, 194].

Extension of confluent hyperintensities into the temporal poles is rare and may indicate that the pathology is not of conventional vascular origin. If the patient has a positive family history of dementing illness and is known to have suffered previously from strokes and/or migraines, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) should also be considered and genetic testing may be appropriate [69, 195]. A range of other single gene disorders can, however, cause SVD [196].

Perivascular Spaces

perivascular spaces (PVS) are interstitial fluid filled cavities surrounding the cerebral vessel walls and separating them from cortical tissue. Connected to the meningeal interstitial spaces, they act as a prelymphatic drainage system for removal of substances that cannot be transported in the blood or catabolised intracellularly [197]. Lacking any valve mechanisms, the movement of fluid within the PVS channels is reliant on regular pulsations from the adjacent vessel. It has therefore been speculated that reduced blood flow due to vascular disease could impair PVS function, causing them to dilate. Furthermore, prolonged or permanent collection of cellular waste in PVS could lead to an inflammatory response and contribute to a number brain diseases, including the neurodegenerative dementias [197].

PVS are typically microscopic and not visible on conventional MRI, however, larger PVS are visible on T2-weighted and FLAIR sequences. These appear linear when imaged parallel to the course of the vessel, and round or ovoid, with a diameter generally smaller than 3 mm, when imaged perpendicular to the course of the vessel [115]. The appearance of PVS is increasingly prevalent with age and some studies have demonstrated a correlation with other SVD features and worsening cognition [198, 199]. However, PVS have been observed at all ages in the healthy population, therefore, their clinical significance remains uncertain [200]. A new rating scale has recently been published to help study the implications of PVS systematically and in more detail, reflecting a growing interest in this topic [201].

Cerebral Microbleeds

Sporadic cerebral amyloid angiopathy is characterised pathologically by the deposition of Aβ in the walls of cerebral blood vessels. This can lead to vessel wall rupture and haemorrhage, microbleeds, capillary occlusion, blood flow disturbances and microinfarcts [176].

When placed in an MRI scanner, the paramagnetic blood break-down products associated with microbleeds cause distortions in the local magnetic field [202]. This distortion leads to faster dephasing of the MRI signal in these regions and results in small, round hypointensities on gradient echo based pulse sequences (T2*-weighted or susceptibility-weighted if available) [115]. The additional dephasing of peripheral tissue causes a blooming affect around the lesion, causing it to appear larger on imaging than when measured in tissue. Faster dephasing, or magnetic susceptibility effects, increase in proportion to the magnetic field strength, therefore, microbleeds may appear larger at higher field strengths, making them easier to detect [71]. For this reason size cut-points have not been recommended for the identification of microbleeds [71]. Other acquisition parameters that enhance MB detection are longer echo times and higher spatial resolution [203, 204].

microbleeds associated with CAA are more commonly found in lobar regions, while microbleeds associated with hypertension are more frequently found in deep brain regions (basal ganglia, thalamus and brainstem), however, these conditions often coexist in the elderly population [205]. A conservative estimate from a large population based study suggests the incidence of microbleeds in the general population is approximately 10% [206]. T2* hypointensities may also result from calcification, iron deposits (from causes other than microbleeds), haemorrhagic metastasis or diffuse axonal injury. Care should be taken to exclude these mimics and other MR artefacts such as flow voids or signals from temporal bones, when reporting the extent of microbleeds in the brain [204].

3.3.3 A Structured Approach to Assessing Signal Change

Expanding on the general framework presented in Figure 3.1, current evidence from the literature and relevant consensus guidelines (discussed above) are synthesised in Figure 3.5 to provide a practical guide to imaging features sufficient to cause vascular cognitive impairment. Where signal change does not meet criteria, visual rating scales are recommended to quantify vascular burden. While radiology expertise is essential to interpret all possible sources of MR signal change, this algorithm may be useful to guide interpretation within the context of vascular cognitive impairment as defined by current consensus criteria.

3.4 Cerebral Atrophy

Atrophy, reflecting neuronal loss, is a downstream effect of several neurodegenerative processes, and can be assessed in detail with structural MRI. Despite a degree of overlap, the pattern of tissue loss broadly relates to clinical phenotype, and in the small number of cases were pathological or genetic confirmation is available, has also been shown to have predictive value for underlying molecular pathology. Atrophy is best identified on T1-weighted images and assessed on a combination of axial, sagittal and coronal views. As outlined in Section 2.2.1, images should be assessed for asymmetric patterns of atrophy, an anterior-posterior atrophy gradient, or disproportionate localised atrophy within particular brain regions, taking into account the patients age. Specific features are discussed below in terms of brain location.

3.4.1 Frontal Lobe Atrophy

Disproportionate frontal lobe atrophy (see Figure 3.6) is typically associated with behavioural changes, or language difficulties if the dominant left hemisphere is affected, and probable clinical syndromes include bvFTD, PNFA progressive non-fluent aphasia (PNFA), and less frequently, CBS CBS. There is a high probability of an underlying FTLD pathology, although on an individual patient basis it may be difficult to pinpoint the particular molecular signature.

Patients diagnosed with bvFTD have been reported to have symmetrical or asymmetri-

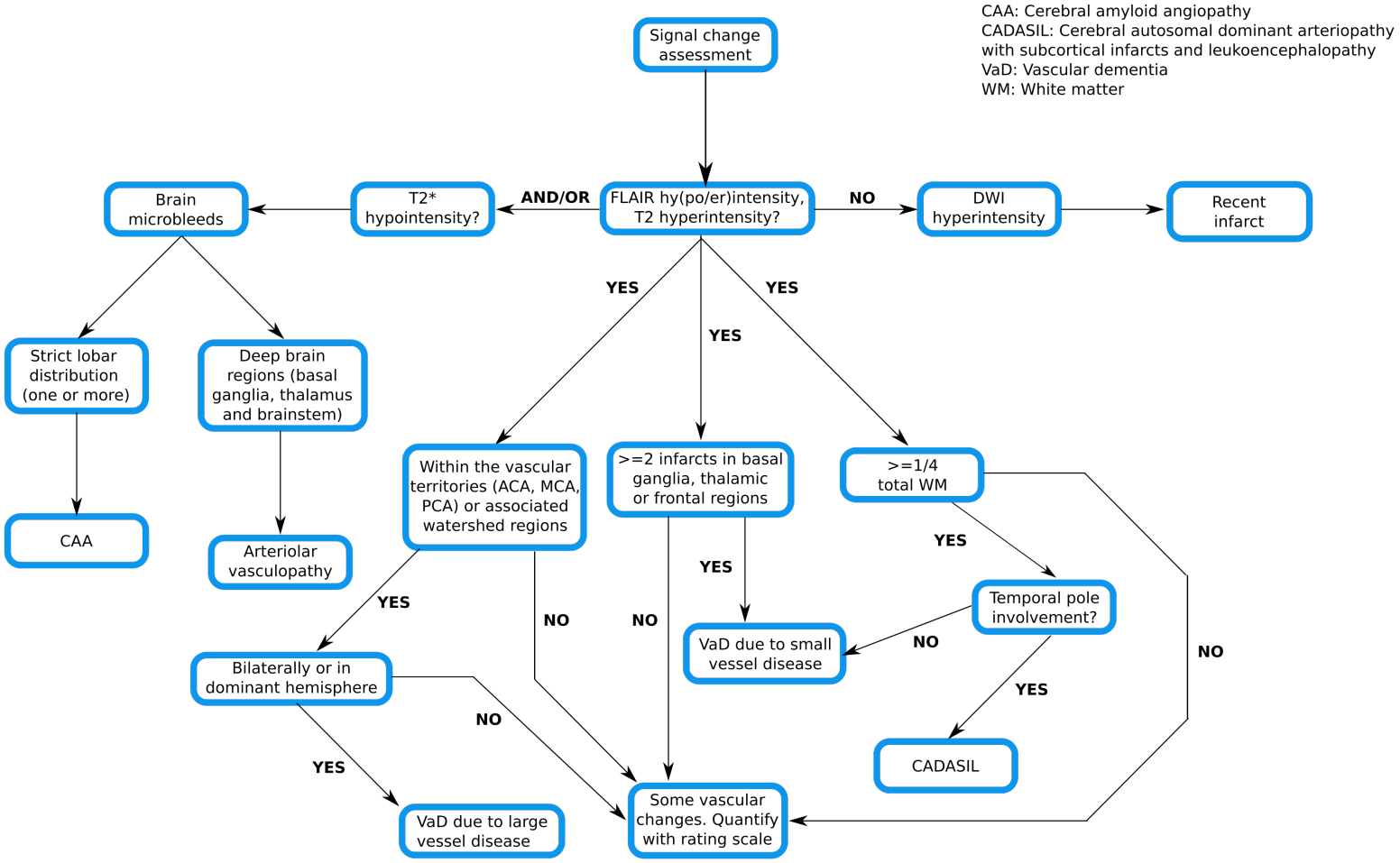


Figure 3.5: An algorithmic approach to assessing MR signal change. This algorithm is intended to guide interpretation within the context of vascular cognitive impairment as defined by current consensus criteria. Radiology expertise is essential to rule out all other sources of MR signal change.

cal frontal atrophy with or without additional temporal lobe atrophy and these findings form part of the criteria for bvFTD [59]. Medial frontal regions,the orbitofrontal cortex and the insula in particular, have been consistently reported in association with bvFTD [145, 207–212]. In patients with PNFA, left-sided posterior frontoinsular atrophy, which may be limited to a subtle widening of the left sylvian fissure, has been frequently reported [213–215] and has also been included in the corresponding diagnostic criteria [60].

Both Tau and TDP43 pathologies have been associated with frontal lobe atrophy, however, relatively few imaging studies have looked at specific subtypes of Tau or TDP43 pathologies. Figure 3.8 demonstrates the studies providing evidence of predominant frontal atrophy by primary molecular pathology.

Frontal lobe atrophy has been reported in several studies of Tau-CBD [52, 110, 214, 216–223], although there is almost certainly an overlap of cases included in these studies (more than half were based on the Mayo clinic pathology database [214, 217–220, 222]). Despite a typically asymmetric clinical presentation in patients with CBS, only two of eleven studies reported an asymmetric pattern of atrophy in the frontal lobes, with no one hemisphere consistently more affected [216, 219]. However, many patients with a clinical diagnosis of CBS are found to have pathology other than CBD at post mortem [218, 220, 224].

While there have been very few studies of Tau-Picks, all have reported predominant frontal lobe atrophy [110, 221, 222, 225], and three of four have reported asymmetry, with the left hemisphere typically more severely affected than the right [110, 222, 225].

Frontal lobe atrophy has been reported in six studies of Tau-PSP [52, 214, 217, 219, 220, 226], with two reporting additional brainstem atrophy [52, 226], although brainstem atrophy has also been reported in the absence of significant frontal lobe atrophy (described below).

Prominent and bilateral frontal lobe atrophy is the most common feature reported in relation to TDP43-A pathology [222, 227–230], however, asymmetry (and extension into the temporal and parietal lobe) has been noted in cases associated with a progran-

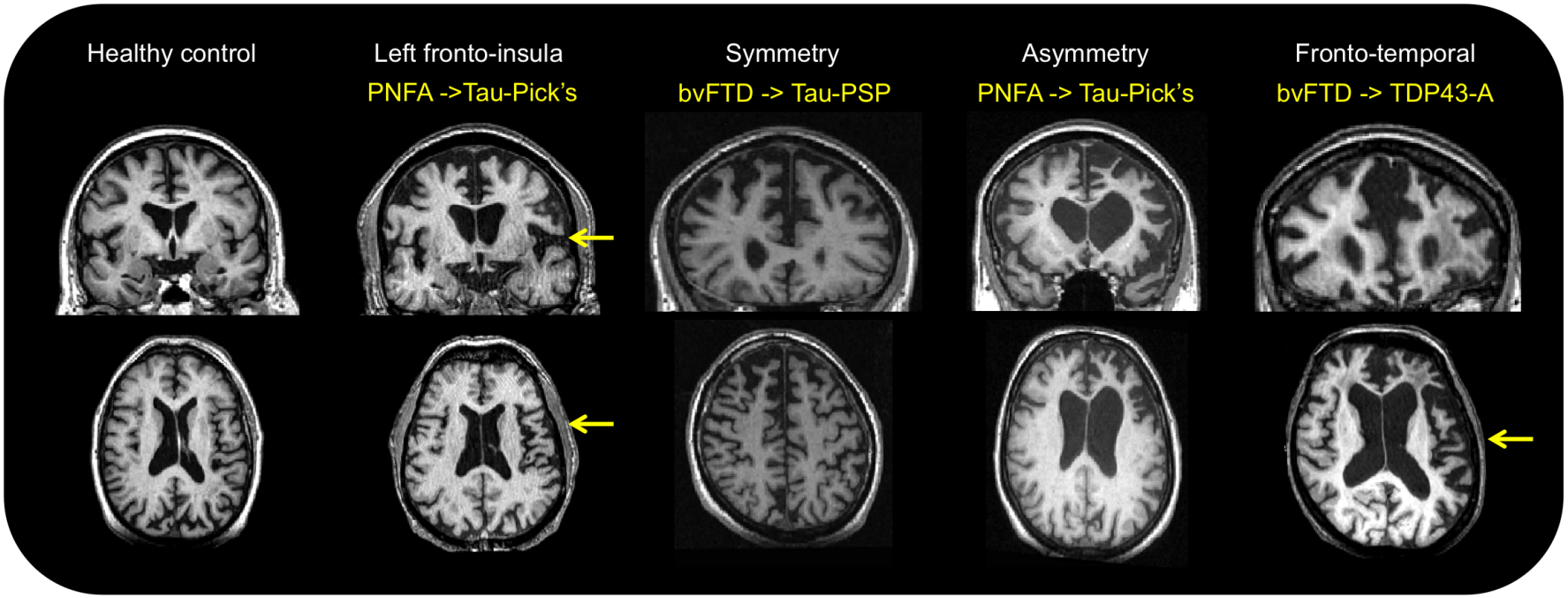


Figure 3.6: Examples of characteristic frontal lobe atrophy in post mortem confirmed cases. The yellow text to the left of the arrow indicates clinical diagnosis, followed by the pathological diagnosis to the right. From left to right: 1) an example of a healthy adult brain, 2) atrophy of the left fronto-insular cortex in a patient with PNFA caused by Tau-Pick's pathology, 3) symmetrical frontal lobe atrophy with Tau-PSP pathology, 4) asymmetric (left>right) frontal lobe atrophy due to Tau-Picks, 5) frontal atrophy extending into the temporal lobes due to TDP43A pathology.

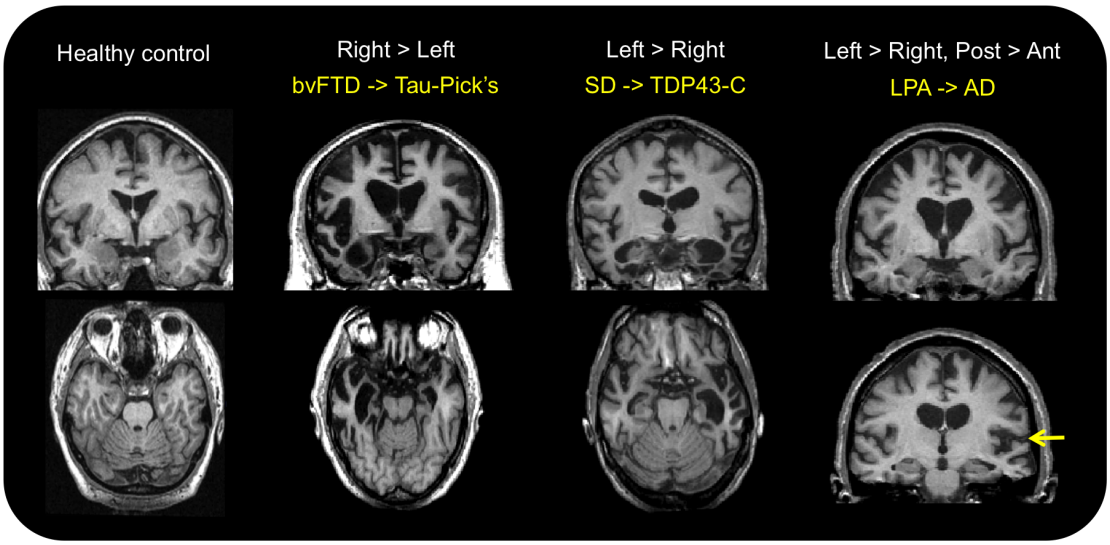


Figure 3.7: Examples of characteristic temporal lobe atrophy in post mortem confirmed cases. The yellow text to the left of the arrow indicates clinical diagnosis, followed by the pathological diagnosis to the right. From left to right: 1) an example of a healthy adult brain, 2) right greater than left sided atrophy in a patient with bvFTD due to Tau-Picks pathology, 3) left greater than right sided atrophy in a patient with semantic dementia and underlying TDP43C pathology, 4) left-sided temporoparietal atrophy in a patient with LPA due to AD

ulin mutation [227]. There have been very few studies of TDP43-B pathology [228, 229, 231], with two reporting predominant frontal lobe atrophy [228, 229], and only one with asymmetry (affecting the left more than the right) [228].

3.4.2 Temporal Lobe Atrophy

Medial temporal lobe atrophy, particularly in the hippocampi, is typically associated with memory problems, while lateral and anteroposterior atrophy (shown in Figure 3.7) is more commonly associated with language difficulties and behavioural changes. An attempt is made to distinguish between predominant medial temporal lobe atrophy, described below as focal hippocampal, and lateral and anteroposterior temporal lobe atrophy (which may also affect the hippocampi) described in this section. Probable clinical symptoms of lateral and anteroposterior atrophy include semantic dementia, logopaenic aphasia and bvFTD, and there is a high probability of underlying FTLD pathology, particularly if there is an asymmetric appearance, although logopaenic aphasia (described below) has been associated with underlying Alzheimer's disease pathology.

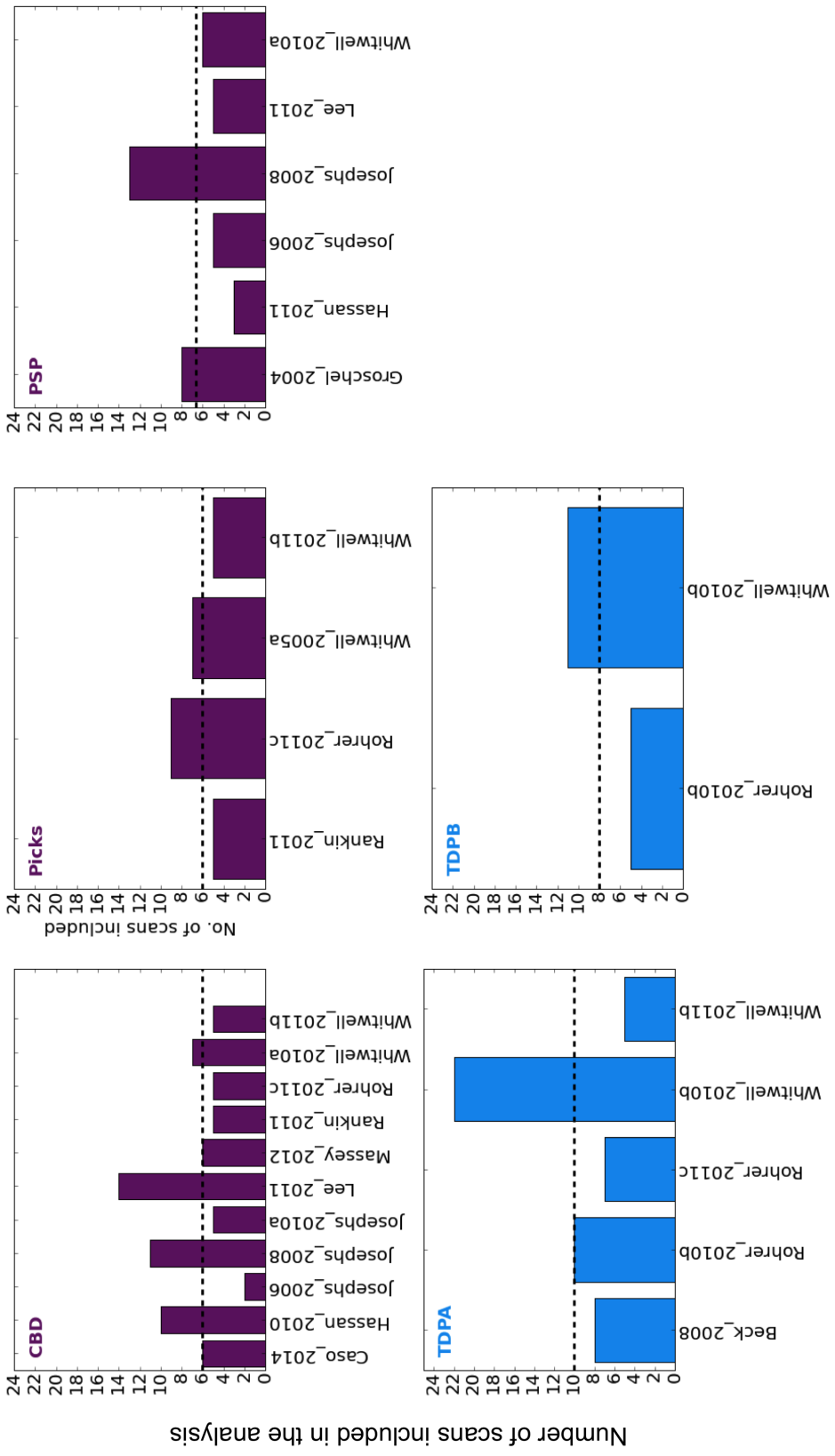


Figure 3.8: Frontal lobe volume loss reported in studies of pathologically confirmed cases. Mean number of cases across all studies is shown by the dotted line.

TDP43-C pathology almost always results in a clinical presentation of semantic dementia and left greater than right sided, focal anterior temporal atrophy, particularly affecting the temporal pole, the amygdala and anterior hippocampus, with selective loss of anterior fusiform gyrus and relative preservation of more posterior structures [112, 208, 227–230, 232]. These imaging characteristics form a core part of the diagnostic guidelines for semantic dementia [60]. Rarely, it results in a behavioural presentation, or with right greater than left sided atrophy [232, 233], with one study suggesting this is linked to additional corticospinal tract degeneration [232]. Imaging studies performed in TDP43-C pathology cases are summarised in Figure 3.9.

Right greater than left sided temporal lobe atrophy is often associated with bvFTD, and often with additional features of prosopagnosia and/or topographical memory impairment, has also been reported in clinical imaging studies [234]. Definitive evidence is lacking to confirm the most common underlying pathology associated with this imaging presentation, however, FTD-MND with behavioural symptoms and underlying TDP43-B pathology has been described in a limited number of cases [231, 235]. While the asymmetry is often striking in semantic dementia and the right temporal variant of bvFTD, both temporal lobes usually become involved and over time the pattern of atrophy becomes more symmetrical [59, 208, 236, 237].

logopaenic aphasia is also associated with greater left-sided atrophy in the temporal lobe although in contrast to semantic dementia the pattern of atrophy extends more posteriorly, predominantly affecting the posterior perisylvian and temporoparietal areas (angular gyrus, posterior middle temporal gyrus, superior temporal gyrus and superior temporal sulcus) [60, 238]. Unlike the majority of clinical syndromes associated with an asymmetrical pattern of atrophy, logopaenic aphasia is typically due to underlying Alzheimer's disease pathology [217, 239–241] (see Figure 3.9).

3.4.3 Focal Hippocampal Atrophy

Hippocampal atrophy, visible with MRI (see Figure 3.10), has been shown to correlate with NFTs burden [242–247]. It is the most established imaging biomarker of Alzheimer's disease and is incorporated in to diagnostic criteria as a supporting feature [37, 39]. Relative preservation of the medial temporal lobe is suggested as a means

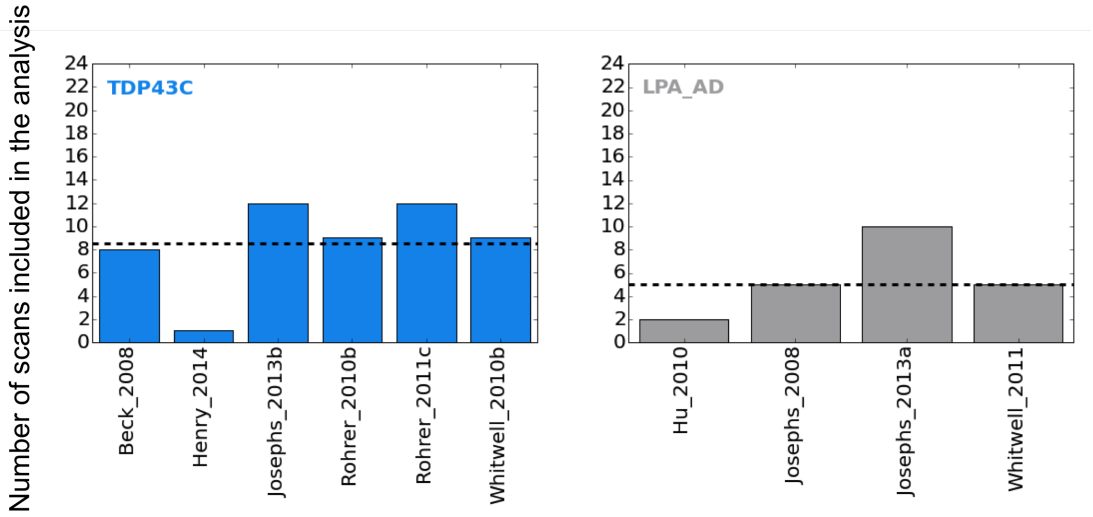


Figure 3.9: Temporal lobe volume loss reported in studies of pathologically confirmed cases. Mean number of cases across all studies is shown by the dotted line.

of distinguishing DLB from Alzheimer's disease at a group level [244, 248–251] (evidence summarised in Figure 3.11), however, the extent to which this is reliable in individual cases is uncertain, not least as the two pathologies frequently coexist [26, 252].

Hippocampal atrophy is also a feature of hippocampal sclerosis and certain inflammatory disorders, hyperintensity of the hippocampus on T2 or FLAIR images makes this diagnosis more likely [253]. Focal (and often severe) atrophy affecting the anteromedial temporal lobes has been described in MAPT mutation carriers (see Figure 3.10), with striking loss of the amygdala, parahippocampus and hippocampal heads bilaterally [212, 225, 227, 228, 230, 254–256], although a few studies have also reported left sided asymmetry [212, 225] (evidence summarised in Figure 3.11). A recent study by Rohrer *et al* in presymptomatic mutation carriers suggests that volume changes in these regions may be detectable up to 15 years before disease onset [256].

3.4.4 Parietal/Occipital Atrophy

Posterior brain atrophy (demonstrated in Figure 3.12), in the region of the parietal and occipital cortex is typically associated with visuospatial and visuoperceptual problems and the posterior cortical atrophy clinical phenotype [257]. Less frequently posterior atrophy has also been demonstrated in patients diagnosed with CBS and DLB [220,

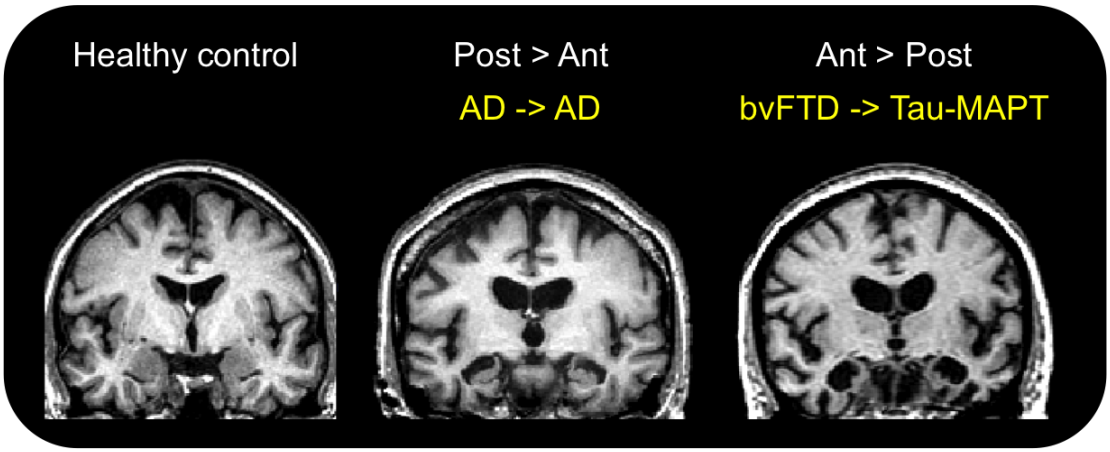


Figure 3.10: Examples of characteristic hippocampal atrophy in post mortem confirmed cases. The yellow text to the left of the arrow indicates clinical diagnosis, followed by the pathological diagnosis to the right. From left to right: 1) an example of a healthy adult brain, 2) bilateral hippocampal in a patient with AD, 3) severe bilateral hippocampal atrophy in a MAPT mutation carrier

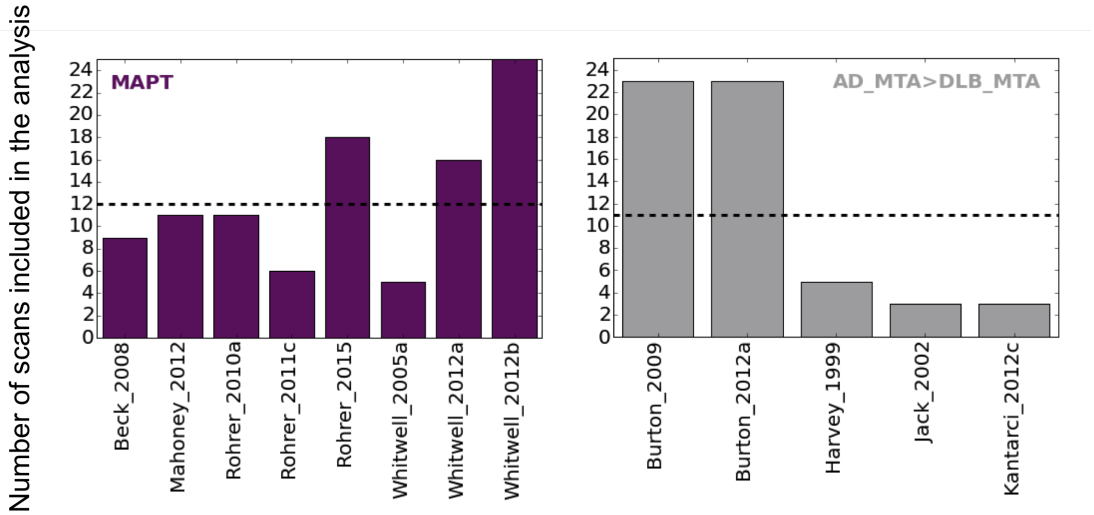


Figure 3.11: Hippocampal volume loss reported in studies of MAPT mutation carriers on the left. On the right, studies providing evidence that MTL atrophy is absent in DLB or less severe than in AD. Mean number of cases across all studies is shown by the dotted line.

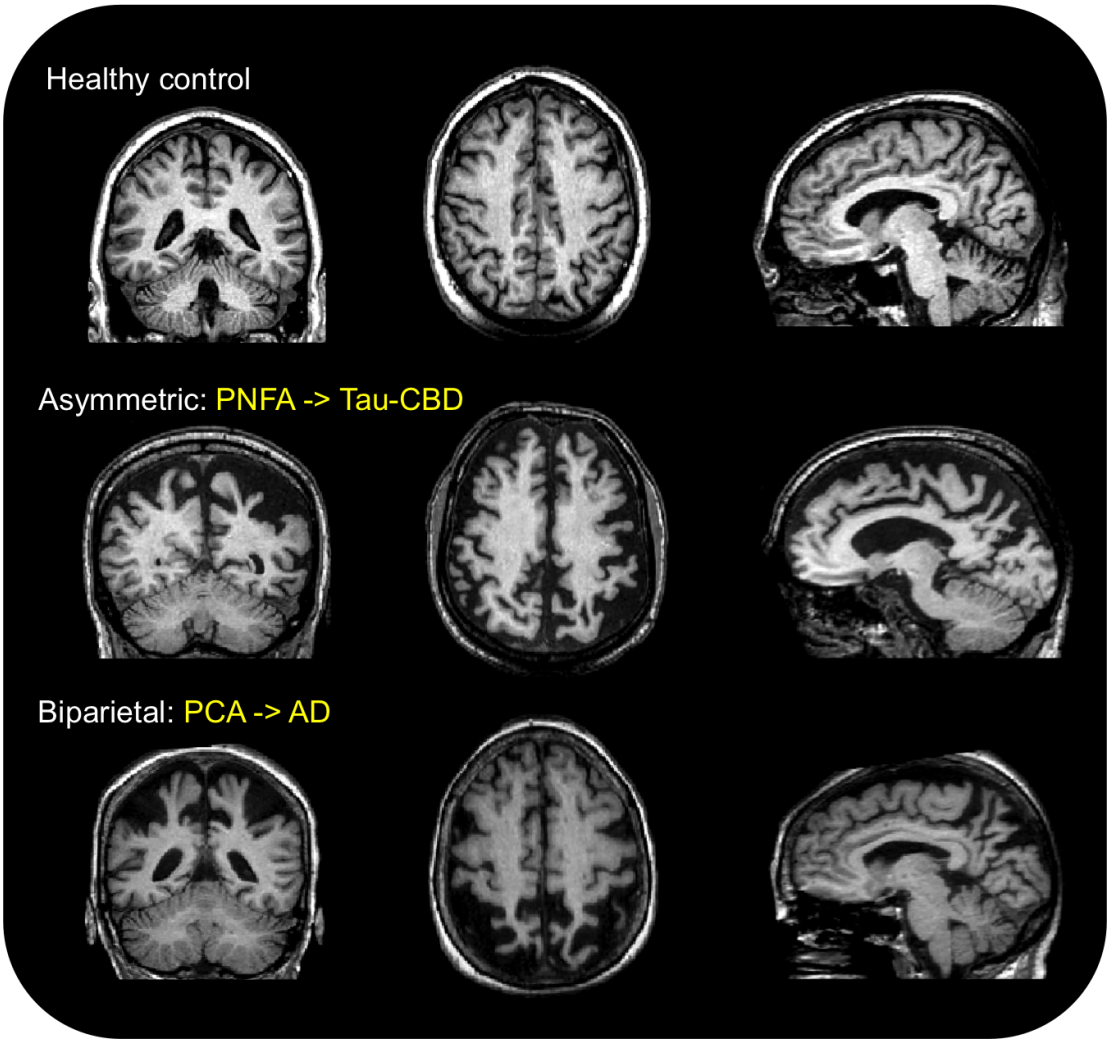


Figure 3.12: Examples of characteristic posterior brain atrophy in post mortem confirmed cases. The yellow text to the left of the arrow indicates clinical diagnosis, followed by the pathological diagnosis to the right. From top to bottom: 1) an example of a healthy adult brain, 2) slightly asymmetric parietal lobe atrophy due to Tau-CBD pathology, 3) severe bilateral parietal lobe atrophy due to AD

258, 259].

Alzheimer's disease is the most common underlying pathology associated with disproportionate posterior brain atrophy [52, 218, 260] . In these patients the medial temporal lobes may be relatively spared, particularly in early-onset cases [111, 139], however, posterior atrophy with additional medial temporal lobe atrophy improves the sensitivity and specificity for underlying Alzheimer's disease pathology [109, 111, 139]. Furthermore, these regions, specifically including the entorhinal cortex, precuneus and posterior cingulate, have also been shown to be affected even in presymptomatic mutation

carriers [101, 261–264].

Tau-CBD has also been associated with parietal atrophy, although less frequently than with frontal lobe atrophy, and without asymmetry [220, 221, 223, 265]. Atrophy associated with GRN mutation carriers is often reported as asymmetric, affecting the entire hemisphere, including frontotemporal and parietal regions [227, 228, 255]. No pathologically confirmed imaging studies could be found to verify evidence of atrophy in posterior brain regions in DLB, however, occipital posterior parietotemporal hypometabolism has been reported [266].

3.4.5 Infratentorial Atrophy

Midbrain atrophy is characteristic of Tau-PSP [52, 212, 223, 226, 267, 268], and has been described in clinical imaging studies as having a hummingbird [269] or 'morning glory' [270] appearance in midsagittal slices, with axial slices demonstrating the so called mickey mouse sign [271] (see Figure 3.13). Simple linear measurements in the midbrain and pons, and calculation of the midbrain to pons ratio, may be a more reliable means of identifying Tau-PSP (ratio<0.5), with a sensitivity of 79% and a specificity of 100% in a sample of cases with pathologically confirmation [268]. Other imaging features in PSP include dilation of the third ventricle and atrophy of the red nucleus [223, 272].

Pathologically confirmed DLB has also been reported in association with brainstem atrophy [267], particularly in the dorsal mesopontine region [244].

3.4.6 A Structured Approach to Assessing Cerebral Atrophy

Current evidence from the literature and relevant consensus guidelines (discussed above) are synthesised in Figure 3.14 to help guide radiological assessment. Since no imaging feature is 100% sensitive and specific, this approach is not intended to be definitive, aiming however, to narrow the differential diagnosis and provide a framework for image assessment and reporting.

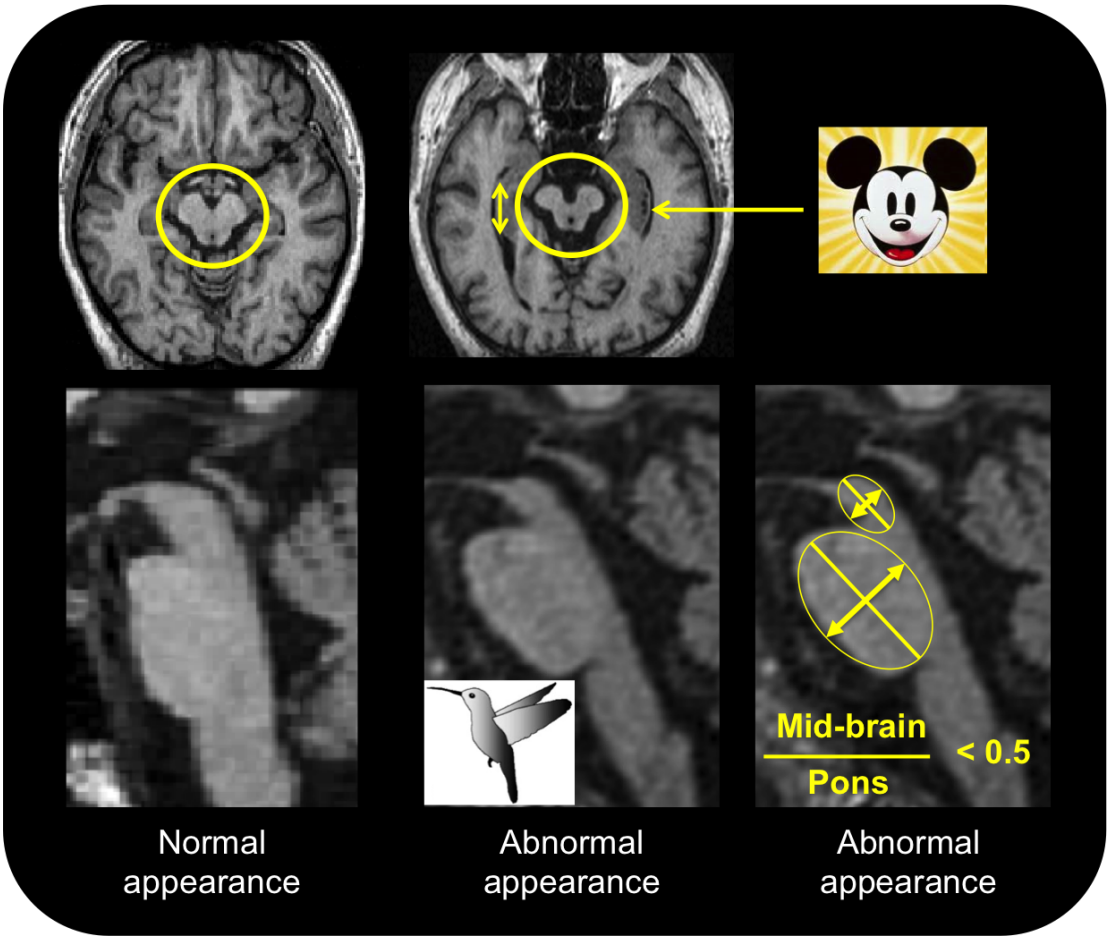


Figure 3.13: Examples of characteristic brainstem atrophy in clinically diagnosed PSP. The top row demonstrates a normal brainstem appearance on the axial slice on the left. The slice on the right demonstrates a reduction in the anterior-posterior diameter of the midbrain resulting in the so-called 'Mickey Mouse' sign. On the bottom row sagittal slices through the brainstem demonstrate a normal appearance (left) followed by an abnormal 'hummingbird' appearance (middle). The image on the right demonstrates the midbrain to pons ratio measurements that have proved useful for identifying Tau-PSP [268]

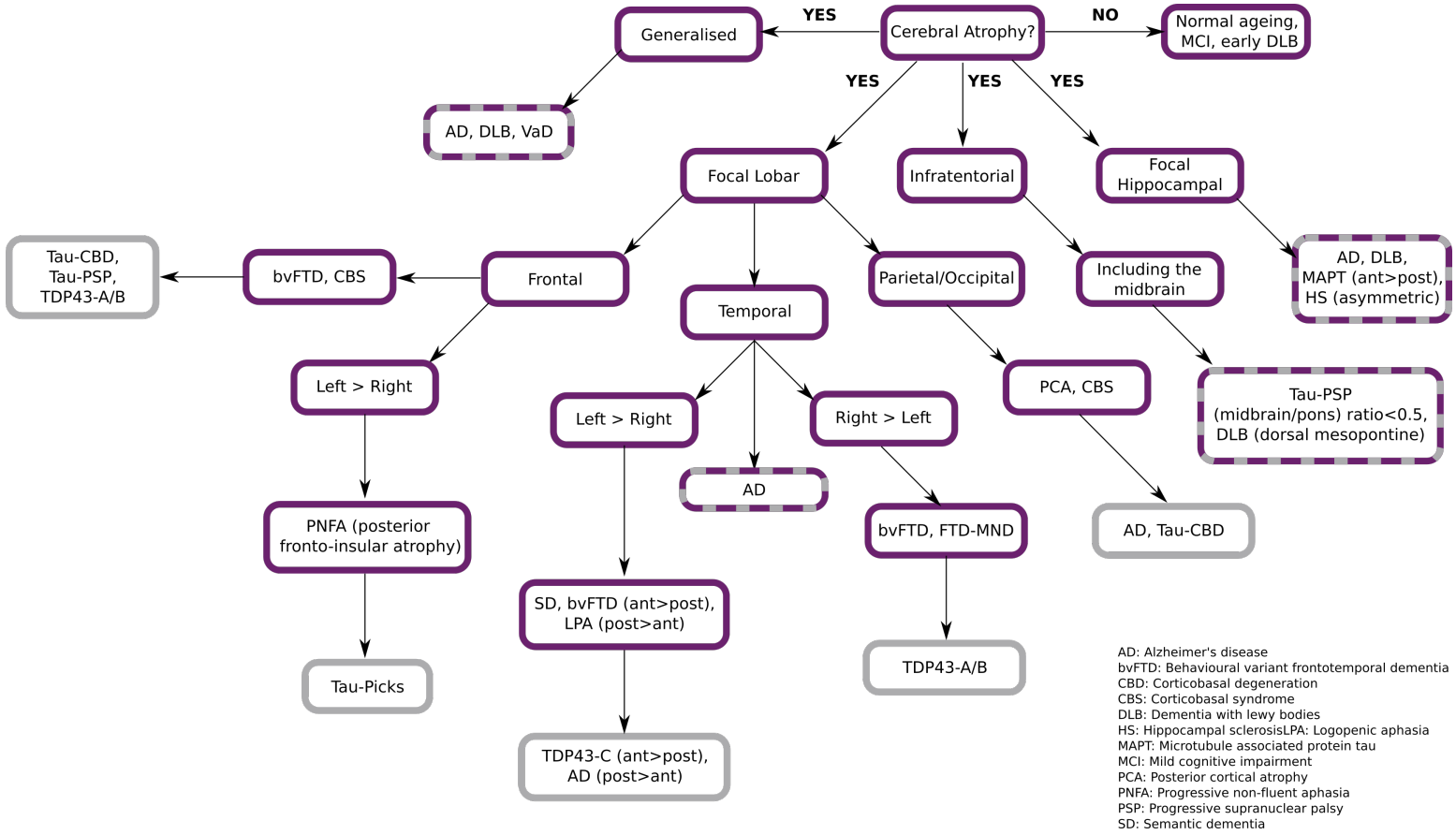


Figure 3.14: An algorithmic approach to assessing MR signal change. The grey boxes indicate the most likely underlying pathology based on the current limited evidence in the literature. The dashed grey and purple boxes represent both the clinical syndrome and underlying pathology.

3.5 No Imaging Abnormality Demonstrated

Following the general framework presented in Figure 3.1, if all stages fail to reveal any abnormalities, beyond what is expected for age, the scan is reported as within normal limits, however, this does not rule out neurodegenerative dementia entirely. If clinical suspicion persists, a SPECT or PET scan can be helpful to look for changes in cerebral perfusion/metabolism, or repeat MRI scanning in six to twelve months. If available, molecular biomarkers such as CSF or amyloid PET scanning are also useful to exclude the presence of amyloid pathology in these cases.

3.6 Conclusion

Structural MR imaging is recommended in the investigation of all neurodegenerative dementias. It is widely available and as demonstrated above, can serve multiple purposes, including the exclusion of space occupying lesions, establishing the extent of vascular disease, and the identification of cerebral atrophy. While each step is equally important in the diagnostic process, the focus of this thesis is explore opportunities to enhance the last of these steps.

As described above, focal and general patterns of atrophy can help to narrow the differential diagnosis, going beyond clinical syndrome to provide an indication of molecular pathology. Several studies of pre-symptomatic genetic mutation carriers have demonstrated that despite being a downstream effect of the disease process, atrophy is often evident before the onset of clinical symptoms. While this has important implications for patient care, it can also be used as part of the enrichment strategy for clinical trials, whilst simultaneously providing a surrogate marker of disease progression and a useful secondary measure to supplement cognitive endpoints.

There is clear motivation to make better use of the valuable, and widely available, diagnostic information provided by structural MRI, for both clinical and research purposes. However, the majority of published studies are conducted in relation to clinically diagnosed cases, which may be wrong in around 30% of cases [3], using a variety of techniques, and often based on small sample sizes. As a result, findings are frequently inconsistent, making it challenging to operationalise them for wider application.

Very few imaging studies have been carried out in relation to pathologically confirmed dementias due to the scarcity of the data. As discussed above, and demonstrated in various figures, the few that are available tend to be based on very small sample sizes (mean n =10). In an attempt to overcome these shortcomings, many studies are forced to pool pathologies together, e.g. as a mixed FTLD group containing both Tau and TDP43, however, valuable information may be lost by combining the data in this way. Furthermore, the pool of available pathologically confirmed data may be even smaller than it appears, as most of it originates from only a few specialist centres. The same images are, therefore, likely to be used to inform multiple studies, introducing a degree of circularity.

Imaging-genetics studies such as the Dominantly Inherited Alzheimer's Network (DIAN), or the GENetic Frontotemporal dementia Initiative, have recognised that to achieve meaningful sample sizes, multiple groups must come together to share valuable resources. However, the same data sharing model has been lacking in terms of pathologically confirmed cases (although the ADNI has now started to acquire pathological confirmation in a few cases). The work presented in this thesis, based on retrospective data collected in three European centres over a twenty year period, is an attempt to pool valuable clinical, imaging and pathology data in relation to multiple different neurodegenerative dementias. A variety of techniques are applied in an attempt to gain insight into the patterns of atrophy associated with various neurodegenrative pathologies, with an interest in translating this information back into clinical diagnostics. In particular this thesis aims to tackle the following problems:

1. To distil characteristic imaging features reported in the literature into a clinically useful algorithm that can be used to interrogate structural imaging, and to test its efficacy against pathological diagnosis.

2. To investigate the diagnostic value of visual rating scales widely used in dementia research, and the feasibility of applying them in a clinical setting.

3. To provide a comprehensive estimate of the global patterns of atrophy associated with different dementia pathologies, and to look for differential patterns that may help to guide visual assessment tools and region of interest analysis.

4. To assess regional differences in brain volume that can help to distinguish between different dementia pathologies.

5. To determine the additional diagnostic value of combining standard measures of assessment in an automated multivariate classifier.

6. To build a sustainable platform to allow this valuable dataset to grow, and to look for opportunities to share this unique and valuable dataset with the wider dementia research community.

Chapter 4

Materials and Methods

4.1 Introduction

Motivated to expand on the existing literature describing cerebral volume loss in typically small samples of gold standard pathologically proven neurodegenerative dementias (described in Section 3.4), a multi-centre, international collaboration was formed between specialist cognitive clinics and research facilities in London, Newcastle and Amsterdam. The work of this thesis has included: the practical set up of this collaboration, which included cleaning and collating the data, building an online database application to provide access, and analysis of the data. This chapter provides a general overview of the data gathering process and the materials and methods common to the experimental analysis, while the following five chapters describe the analysis and database. Experiment specific details will be provided in the relevant chapters.

4.2 Study Population

All cases included in this study had a diagnosis of dementia during life and post mortem confirmation of underlying pathology. Ethical approval for the study was obtained from the National Research Ethics Service Committee London South East.

4.2.1 Data Collection Sites

The patient data used in this thesis was collected from specialist cognitive clinics and research facilities in London, Newcastle and Amsterdam. All data was transferred to the UCL Dementia Research Centre for processing and analysis.

The majority of patients attended cognitive clinic at the National Hospital for Neurology and Neurosurgery in London or attended for research at the UCL Dementia Research Centre. Some of the older cases attended clinic at St Mary's Hospital in London.

These cases were examined pathologically at the Queen Square Brain Bank for Neurological Disorders or the MRC London Neurodegenerative Diseases Brain Bank.

In Newcastle, data was donated by the Dementia and Memory Services at the Campus for Ageing and Vitality, Newcastle University. Pathological assessment was performed at the Newcastle Brain Tissue Resource (formerly Newcastle Brain Bank).

In Amsterdam, patients attended clinic at the Alzheimer Centre, VU University Medical Centre (VUmc). Pathological assessment was performed at the VUmc Department of Pathology.

The UK based pathological assessment centres are part of the UK Brain Banks Network.

4.2.2 Patient Diagnosis

In total the three centres held data on 511 patients with pathological confirmation of their diagnosis. Of these, imaging data was available for 319 patients, however, only 233 of these were found to have 3D T1-weighted imaging of sufficiently high quality to be used for image analysis. Reasons for exclusion included no 3D scan, gross patient movement, scanner artefact, low tissue contrast, and evidence of other non-neurodegenerative pathologies. Patients who did not have a primary diagnosis of Alzheimer's disease, DLB or FTLD pathology were also excluded (n=49) from this analysis. Examples of excluded diagnosis included prion disease cases, vascular disease in the absence of neurodegenerative pathology, multiple system atrophy, and Parkinson's disease.

In total therefore, 184 patients were identified with histopathologically confirmed dementia: 101 had a primary pathology diagnosis of Alzheimers disease, 28 with DLB, and 55 with FTLD (24 tauopathies, 25 TDP43 proteinopathies, 3 with FUS proteins and 3 with other designations [2 *C9ORF72* mutation carriers, 1 FTLD-MND]). 25 patients also had confirmation of their genetic status: 3 with *APP*, 9 *PSEN1*, 5 *C9ORF72*, 3 *MAPT*, and 5 *GRN*. Pathological examination of brain tissue was carried out between 1997 and 2013 according to standard histopathological processes and criteria in use at the time of assessment. The quality of neuropathological assessment may have varied

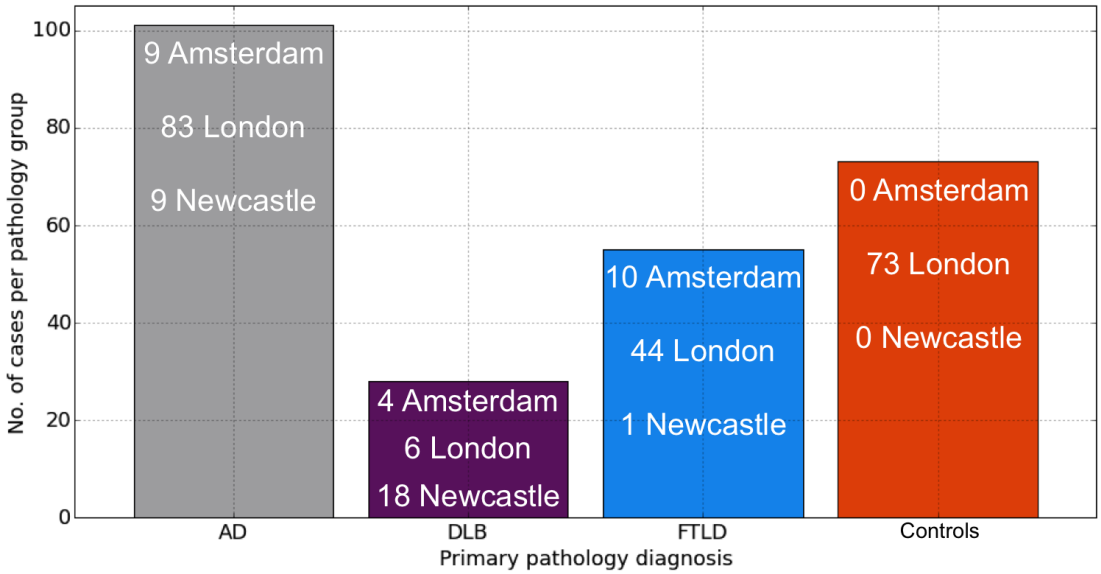


Figure 4.1: Primary pathology groups included in this thesis. The number of images contributed per centre per pathology are listed on each bar above.

over this time however, older FTLD cases from the London sites, previously diagnosed as ubiquitin-positive, tau-negative, were reassessed with TDP43 or FUS immunohistochemistry after the discovery that many patients could be further characterised by these molecular pathologies. Data on secondary pathologies was not reliable in every case, therefore, only primary pathological diagnosis was considered for analysis.

4.2.3 Healthy Controls

73 cognitively normal control participants were also included for analysis, however, with the exception of one participant, pathological confirmation was not available. Control data was contributed exclusively by the UCL Dementia Research Centre from a database of participants who have previously enrolled in research studies. 23 of these subjects were enrolled through the MIRIAD study, the data from which is now available online for research purposes (https://www.ucl.ac.uk/drc/research/miriad-scan-database) [273].

4.3 Structural MRI

The motivation for this thesis was to explore cerebral volume loss in pathologically proven dementias. All analysis is, therefore, described in relation to 3D T1-weighted

imaging, however, any and all available imaging was collected for every subject (including T2-weighted, FLAIR, proton density and diffusion weighted imaging) and could be used for subsequent analysis. As patients and controls were recruited over a twenty year period into various studies, at multiple centres, available imaging was highly variable between subjects, with 3D T1-weighted the most common.

4.3.1 Acquisition

All individuals had T1-weighted volumetric MR imaging performed during life. As the data were collected retrospectively from multiple centres, the images were acquired on scanners from 3 different manufacturers (Philips, GE, Siemens) using a variety of different imaging protocols. Magnetic field strength varied between 1.0T (n=20 scans), 1.5T (n=205 scans) and 3T (n=32 scans).

4.3.2 Image Pre-processing

Imaging data were available in Digital Imaging and Communications in Medicine (DICOM), NIfTI (Neuroimaging Informatics Technology Initiative) or Analyze (developed by the Biomedical Imaging Resource at the Mayo Foundation) format. DICOM images were anonymised and converted to NIfTI format using the MRIcron dcm2nii tool (http://www.mccauslandcenter.sc.edu/mricro/mricron/). Images were converted to DICOM using the dcmbuild tool developed by Dr Mark White of the UCL Medical Physics department. Analyze files were transferred to and from NIfTI using the in-house midas2nii or nii2midas tool. Image files were converted to all three image formats for various tasks:

- **DICOM**: For integration with the online image viewer included in the database web application (Chapter 9).
- **NIfTI**: For VBM analysis and image registration and segmentation using NiftyReg and NiftySeg packages (Chapter 7 and Chapter 8).
- **Analyze**: For INU correction (see below), creating a whole brain mask using the Brain-MAPS algorithm (Chapter 8), and for visual assessment (see below).

4.4 Image Analysis

Details of experiment specific analysis are provided in individual chapters. The techniques below are common among chapters and are described here to avoid repetition.

4.4.1 Intensity Non-Uniformity Correction

All images were corrected for INU prior to experimental analysis. INU correction was performed using the nonparametric nonuniformity normalisation (N3) method [274]. N3 models the INU as a smooth multiplicative field and uses the image intensity histogram to iteratively smooth this field whilst maximising the high frequency component (which is reduced as a result of the image blurring that occurs with INU). N3 does not require any a priori knowledge about the underlying anatomy and is one of the most stable and widely used INU correction methods [143]. This technique is fully automated.

4.4.2 Visual Assessment

The visual assessment and visual rating experiments described in Chapter 5 and Chapter 6 were carried out using the Medical Information Display and Analysis System (MIDAS) viewer [275]. The viewer allows simultaneous multiplanar display of 3D imaging data. Image contrast and zoom can also be adjusted. MIDAS was also used for initial quality control of the images.

4.4.3 Calculating Total Intracranial Volume

Brain volume studies are confounded by the variability in pre-morbid brain volume. In studies of neurodegenerative disease, failure to correct for this variability in the statistical model will diminish its power to accurately detect group differences or correlations between structural volumes and clinical measures. By including an estimate of TIV, which remains constant over time as the CSF expands to fill the space created by cerebral atrophy, this potential problem can be controlled [167].

Manual outlining is considered the gold standard for estimating TIV, however, this is

not practical in large studies. Ridgway *et al* compared several automated methods against manual outlining [276]. In Chapter 6, Chapter 7 and Chapter 8, TIV was calculated based on Jacobian integration of the forward deformation fields generated by the unified segmentation model. This method was shown to be highly correlated with manual outlining (Pearson correlation coefficient=0.97) [276].

4.4.4 Voxel Based Morphometry

In Chapter 6 and Chapter 7, VBM pre-processing and analysis (described in Section 2.3.4) was performed using SPM12b (Statistical Parametric Mapping, Version 12b revision 5829; http://www.fil.ion.ucl.ac.uk/spm) and Matlab version R2012a (7.14.0.739 - 64-bit, uk.mathworks.com/products/matlab/). Due to the variability in scanning parameters an initial rigid registration to the Montreal Neurological Institute ICBMI152 template was performed using the Reg-Aladin tool from the NiftyReg package [277, 278] to provide a better starting point for the SPM pre-processing pipeline. Each registration was then checked and manually adjusted (if necessary) such that the anterior commissure was within a few millimetres of the origin and the orientation was within a few degrees of the ICBMI152 template. Grey matter, white matter and CSF were obtained using the unified segmentation approach [162], which includes bias correction (regularisation=0.001, full-width at half-maximum (FWHM)=60mm) and rigid registration to the ICBMI152 template. A group average tissue probability map was generated through iterative alignment of the initial segmentations to an evolving estimate of their group-wise average using the DARTEL toolbox [163, 164] and affinely registered to the ICBMI152 template. The initial grey and white matter segmentations were then warped using the DARTEL transformations and computed flow fields, modulated to account for local volume changes, then smoothed with a 6mm FWHM Gaussian kernel.

4.5 Statistical Analysis

4.5.1 Python Programming Language

All (non-imaging) data processing and statistical analysis methods were implemented in Python version 2.7.6 (64-bit). Python is a general purpose, open-source programming language widely used by the scientific community and increasingly applied in neuroscientific research [279–281] (recently a dedicated research topic of an issue of Frontiers in Neuroinformatics). It has an extensive collection of libraries for scientific and numeric computing. The main libraries used in this thesis are NumPy (version 1.8.1), SciPy (version 0.14.0), Pandas (version 0.14.1), Matplotlib (version 1.4.3) and Scikit-Learn 0.15.2.

- NumPy [282] and SciPy [283] are core modules for all mathematical and statistical operations.
- Matplotlib is a plotting library and is used to create all of the graphical figures in this thesis [284].
- Pandas is a powerful tool for efficient data manipulation and analysis [285].
- Scikit-Learn provides a wide range of state-of-the-art machine learning algorithms [286].

One of the many benefits of using Python is that entire data analysis workflows can be implemented in a single programming language, allowing them to be more easily managed and replicated.

4.5.2 VBM Analysis

Individual design matrices are discussed in Chapter 6 and Chapter 7, however, in both experiments analysis was restricted to voxels contained within an explicit mask. The mask was created, based on the optimal threshold of the group average image, using the automatic mask creation strategy in the SPM toolbox [165] (see Figure 4.2).

In both experiments, results are overlaid on a group average template. This was created by affinely registering the final DARTEL template to the ICBMI152 template and using

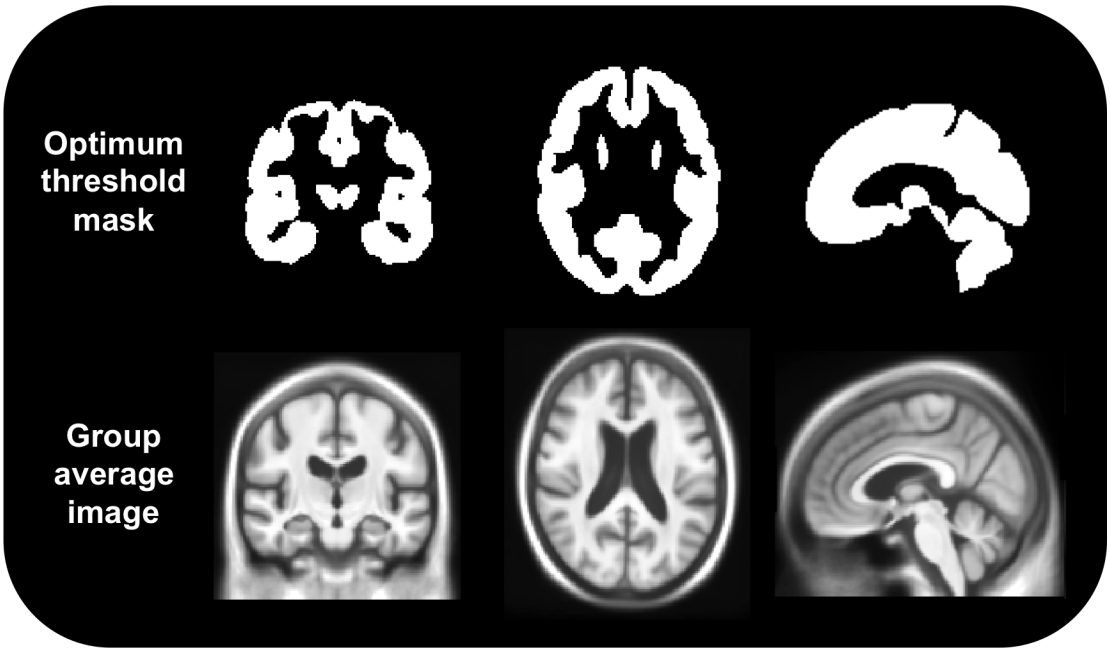


Figure 4.2: VBM mask based on the optimal threshold of the group average image. Group average image based on affinely registering the final DARTEL template to the ICBM152 template and using this transformation and the computed flow fields to warp the bias corrected, whole brain images.

this transformation and the computed flow fields to warp the bias corrected, whole brain images (see Figure 4.2). In this case no modulation or smoothing was applied. The resulting images were then averaged to produce an image template representative of both the study population and the pre-processing steps applied.

4.5.3 Support Vector Machine

Linear SVMs were used for binary classification tasks in Chapter 6 and Chapter 8. In all experiments split half separation of the data was applied, i.e. half of the data was used for training and half reserved for testing purposes. The training data was scaled to zero mean and unit variance over subjects, with the same transformation then applied to the testing data. SVMs were trained using leave-one-out cross-validation on the training data and class weighting was applied to adjust for unbalanced groups. The regularisation parameter, c, was optimised using grid-search in the range 1E-5 to 100, increasing by an order of magnitude with every iteration.

Chapter 5

An Algorithmic Approach to Image Assessment

5.1 Introduction

As outlined in Chapter 3, signal change and cerebral atrophy visible on structural MRI can be used to identify diagnostically relevant imaging features, which provide support for underlying pathology, and help to narrow the differential diagnosis. Currently, both radiology expertise and thorough knowledge of the neuroimaging literature are required in order to identify many of these features. Such expertise is however, unavailable in many clinical settings. By incorporating important imaging characteristics and guidance from relevant consensus criteria in a structured algorithm, it may be possible to make this information more accessible to a wider clinical audience and provide a practical guide to scan assessment in neurodegenerative dementia.

In this chapter, the algorithmic approach to identifying cerebral atrophy, presented in Section 3.4, is tested by five image analysts using imaging acquired during life in patients with a pathology-confirmed diagnosis of dementia and healthy control subjects. The aim of this study was to evaluate the reliability with which analysts could agree on a predominant pattern of atrophy based on unstructured visual assessment of the image; and as an estimate of the utility of visual assessment for differential diagnosis, to determine the frequency with which each pathology group was assigned to each branch of the algorithm.

5.2 Methods

5.2.1 Study Population

The study population is described in Section 4.2. In brief, it consists of 184 patients with a diagnosis of dementia during life and pathological confirmation of underlying pathology: 101 had a primary diagnosis of Alzheimer's disease, 28 with DLB, and 55 with FTLD (24 tauopathies, 28 TDP-43 proteinopathies, 3 fusopathoies). 23 healthy control subjects previously described in the MIRIAD study were also included in the analysis [105, 273].

5.2.2 Structural MRI

Image acquisition is described in Section 4.3. All individuals had 3D T1-weighted volumetric MR imaging acquired during life.

5.2.3 Image Assessment

Assessment of the complete imaging dataset (n=207) was carried out by two analysts (Lorna Harper and Dr Giorgio Fumagalli), both with over three years experience in analysing structural MRI scans in dementia. To provide further independent testing, three additional experienced analysts (Prof. Nick Fox, Prof. John O'Brien and Dr Jonathan Schott), each with over fifteen years experience in analysing structural MRI scans in dementia, also assessed 80 scans (20 Alzheimer's disease, 20 DLB, 20 FTLD (1 FUS, 7 Tau, 12 TDP43), 20 controls) drawn at random from the total study population. Images were viewed using the in-house MIDAS image viewer (described in Section 4.3) [275] and displayed in a random sequence in terms of underlying pathology. Analysts were asked to follow the algorithm outlined in Figure 3.14 to identify the most prominent atrophy pattern on each scan and the single most likely clinical and pathological diagnoses associated with it. Image assessment was carried out blind to all clinical and pathological information except age at the time of scanning. The algorithm choices made by each rater were recorded in individual on-line forms (an example form is included in Appendix C). LH and GF initially performed visual rating training, applying the protocol described above to a sample of 20 images (5 controls,

15 with a clinical diagnosis of dementia) from research participants who attended the Dementia Research Centre, London. Five images were assessed by consensus, then a further fifteen independently, which were then reviewed by consensus.

5.2.4 Inter-Rater Reliability

Inter-rater reliability was calculated at the level shown in Figure 5.1, i.e. the decision between the categories no atrophy, generalised atrophy, predominant frontal, temporal, hippocampal, parietal or infratentorial atrophy. inter-rater reliability was calculated using Fleiss' kappa statistic, which is an extension of Cohen's kappa used to determine agreement between more than two analysts. Kappa statistics are defined as

$$\kappa = \frac{Pa - Pe}{1 - Pe}$$

Where Pa is the proportion of observations in agreement between analysts and Pe is the proportion in agreement due to chance. In terms of Fleiss' kappa, for every scan i=1,..,n and every category j=1,..,k, X_{ij} = the number of analysts that assign category j to scan i, and Pa is defined as

$$Pa = \frac{1}{mn(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} X_{ij}^2 - mn$$

Pe is given by

$$Pe = \frac{1}{(mn)^2} \sum_{j=1}^{k} \left(\sum_{i=1}^{n} X_{ij} \right)^2$$

inter-rater reliability was calculated independently for the two analysts that assessed the complete data set. It was also calculated for the three more experienced analysts based on the subset of scans, and then for all five analysts based on the common subset. Finally, agreement between all unique analyst pairs was calculated based on the common subset to determine the maximum level of agreement achieved using this dataset. In the case of only two analysts, Fleiss' kappa is equivalent to Cohen's kappa. Ninety-five percent confidence interval (CI)s were also calculated for each kappa value (see

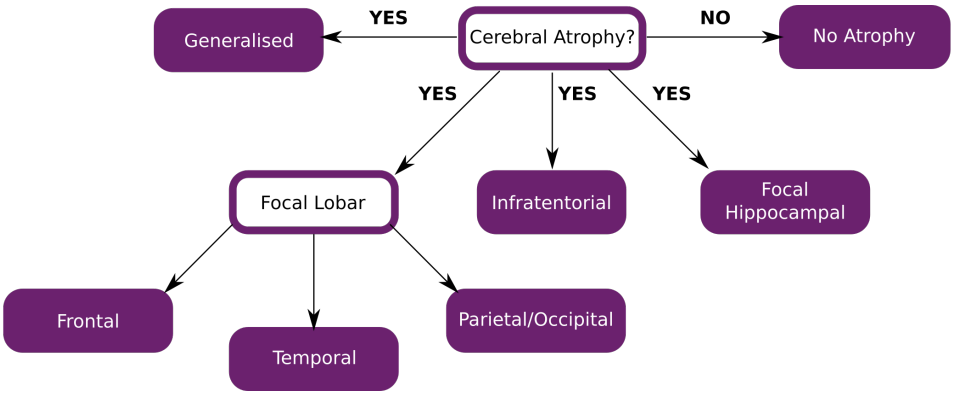


Figure 5.1: Primary decision level used to assess inter-rater reliability is shown by the purple boxes with white text.

Appendix D for details).

5.2.5 Frequency Distribution of Pathology Groups

The relative frequency distribution of pathology groups to each branch of the algorithm was determined individually for each analyst and the results plotted side by side for comparison. The frequency distribution was determined based on the complete dataset (207 images, 2 analysts), and the random subset (80 images, 3 analysts).

All data processing and analysis, including IRR, were performed using Python libraries NumPy 1.8.1, SciPy 0.14.0 and Pandas 0.14.1 on Python 2.7.6 64-bit.

5.3 Results

5.3.1 Demographics

Demographic details of the patients and controls included in the complete group are shown in Table 5.1. The DLB and control groups were significantly older than the TDP43 group ($p<0.05$). The control was also older than the Alzheimer's disease group ($p<0.05$). The average age of the small FUS group was notably younger than the other groups, however, the sample size of this group was too small to perform meaningful statistical analysis. There were no significant differences in disease duration at the time of scanning between any of the pathology groups. The number of years between scan and death was significantly different between the Alzheimer's disease and DLB group

($p<0.05$), with the Alzheimer's disease group living slightly longer on average. Mini-Mental State Examination (MMSE) within six months of scan date was only available in 116 of the 184 patients (missing data in 37/101 Alzheimer's disease, 5/28 DLB, 26/55 FTLD). Based on the data available, there was a significant difference in average MMSE scores between the Alzheimer's disease and Tau group ($p<0.05$), with the Alzheimer's disease group more affected. Very similar results were found when using MMSE closest to scan where data was available for 170 patients. The small FUS group had a notably higher MMSE score than the other groups. Demographic details of the patients and controls included in the subset group are shown in Table 5.2. This data broadly reflects the total study population, although the Alzheimer's disease group taken as a whole were slightly more affected in the subset group, while the FTLD-Tau group were slightly older.

5.3.2 Inter-Rater Reliability

Agreement (i.e. the single region with the greatest atrophy) was reached between the two analysts assessing the complete dataset in 110 of 207 scans (53%), resulting in a κ value of 0.43 (95% CI: 0.37-0.50). A specific breakdown of agreement based on algorithm branch and pathology is demonstrated in Table 5.3. Agreement was highest in the assessment of the control group (17 of 23 scans (74%)), followed by the FTLD-TDP43 group (20 of 28 (71%)) and the small FTLD-FUS group (2 of 3 (67%)). Agreement was lowest among the DLB group (12 of 28 (43%)).

The three experienced analysts reached complete agreement in 32 of 80 scans (40%) in the subset. Agreement between at least two them was reached in 70 of 80 scans (88%). Fleiss' kappa was calculated as κ=0.44 (95% CI: 0.37-0.50). Table 5.4 presents the breakdown of this agreement. Agreement (between two experienced analysts) was reached in 100% of the FTLD-TDP43 cases and the single FTLD-FUS case, 90% of the DLB and control cases, and 80% of the AD cases. The greatest variability was seen in the FTLD-Tau group.

Based on the five analysts assessing the 80-scan subset, complete agreement was reached in 25 of 80 scans (31%). Agreement between four analysts was reached in 40 scans (50%), between three analysts in 59 scans (74%), and between at least two an-

	Control	AD	DLB	Tau	TDP43	FUS
N	23	101	28	24	28	3
Gender (%M)	52	61	75	58	50	100
Age at scan (yrs)	70 (7)	61 (11)	70 (6)	64 (9)	60 (8)	49 (3)
Disease duration at scan (yrs)	-	4 (3)	3 (2)	4 (2)	3 (3)	2 (1)
Time from scan until death (yrs)	-	6 (3)	4 (2)	5 (3)	6 (3)	4 (1)
MMSE (x/30)	-	17 (6)	20 (5)	23 (5)	22 (7)	28 (2)

Table 5.1: Study population demographics. Results are shown as mean (standard deviation). MMSE is within six months of imaging acquisition

	Control	AD	DLB	Tau	TDP43	FUS
N	20	20	20	7	12	1
Gender (%M)	55	50	70	43	50	100
Age at scan (yrs)	69 (8)	61 (11)	71 (6)	70 (7)	58 (9)	46 (NA)
Disease duration at scan (yrs)	-	5 (4)	3 (3)	3 (2)	3 (4)	2 (NA)
Time from scan until death (yrs)	-	4 (2)	3 (2)	5 (3)	6 (3)	5 (NA)
MMSE (x/30)	-	14 (6)	20 (5)	25 (2)	21 (8)	26 (NA)

Table 5.2: Subset study population demographics. Results are shown as mean (standard deviation). MMSE is within six months of imaging acquisition

alysts in 80 scans (100%). Fleiss' kappa was calculated as κ=0.44 (95% CI: 0.41-0.48). Table 5.5 presents the breakdown of agreement between at least three analysts. Agreement was higher among the FTLD-TDP43 and control scans (92% and 85% respectively). Agreement was lowest for the Alzheimer's disease and FTLD-Tau groups (55% and 57% respectively).

Finally, agreement between all unique analyst pairs was calculated. The highest calculated kappa value was between two of the experienced analysts, κ=0.57 (95% CI: 0.46-0.67). Lowest agreement between any two analysts was κ=0.34 (95% CI: 0.23-0.45).

5.3.3 Frequency Evaluation at the Individual User Level

The relative frequency with which each pathology group (in the complete dataset) was assigned to each of the main branches of the decision algorithm (generalised, frontal, temporal, hippocampal, parietal, infratentorial - see Figure 5.1) is presented in Figure 5.2. Asymmetry within the frontal and temporal branches is displayed as hatched regions.

Both analysts classified more than 50% of the Alzheimer's disease cases as having either predominant parietal or hippocampal atrophy, although analyst 2 selected parietal lobe atrophy more frequently. Analyst 1 also classified around a fifth of cases with temporal lobe atrophy, while analyst 2 assigned roughly the same proportion to generalised atrophy. Similarly in the DLB cases, analyst 2 interpreted a generalised atrophy pattern more often, while analyst 1 distributed the cases more evenly among parietal, generalised and no atrophy, and a slightly higher proportion with temporal predominant atrophy. Both analysts classified a high proportion (>50%) of FTLD-Tau cases with predominant frontal or temporal lobe atrophy, and often with left-sided asymmetry. Left-sided temporal lobe atrophy was also the dominant finding in patients with FTLD-TDP43 for both analysts. Both analysts also classified a high proportion of controls as having no atrophy.

Figure 5.3 presents the relative frequency with which each pathology group in the imaging subset was assigned to each of the main branches of the decision algorithm by the experienced analysts. Based on the imaging subset analysed, analyst 1 classified the

	No Atrophy	Frontal	Temporal	Hippocampal	Parietal	Infratentorial	Generalised	Total
Controls	16 (70%)	-	-	-	1 (4%)	-	-	17 (74%)
AD	5 (7%)	1 (1%)	2 (3%)	14 (19%)	21 (29%)	-	3 (4%)	46 (63%)
DLB	3 (11%)	-	-	2 (7%)	3 (11%)	-	4 (14%)	12 (43%)
Tau	-	8 (33%)	3 (13%)	-	2 (8%)	-	-	13 (54%)
TDP43	-	3 (11%)	13 (46%)	1 (4%)	1 (4%)	-	2 (7%)	20 (71%)
FUS	-	1 (33%)	-	1 (33%)	-	-	-	2 (67%)
Total	24 (12%)	13 (6%)	18 (9%)	18 (9%)	28 (14%)	0	9 (4%)	110 (53%)

Table 5.3: Algorithm agreement for two analysts assessing the full dataset of scans (n=207). Agreement reached in 110 of 207 scans (53%). Values are shown as count (group proportion)

	No Atrophy	Frontal	Temporal	Hippocampal	Parietal	Infratentorial	Generalised	Total
Controls	17 (85%)	-	-	-	-	-	1 (5%)	18 (90%)
AD	2 (10%)	-	1 (5%)	6 (30%)	3 (15%)	-	4 (20%)	16 (80%)
DLB	7 (35%)	-	-	5 (25%)	2 (10%)	-	4 (20%)	18 (90%)
Tau	-	2 (29%)	2 (29%)	-	-	-	1 (14%)	5 (71%)
TDP43	3 (25%)	1 (8%)	6 (50%)	1 (8%)	-	-	1 (8%)	12 (100%)
FUS	-	-	1 (100%)	-	-	-	-	1 (100%)
Total	29 (36%)	3 (4%)	10 (13%)	12 (15%)	5 (6%)	-	11 (14%)	70 (88%)

Table 5.4: Algorithm agreement among at least two experienced analysts for subset of scans. Agreement was reached in 70 of 80 scans. Complete agreement (among three) was reached in 32 scans (40%). Values are shown as count (group proportion)

	No Atrophy	Frontal	Temporal	Hippocampal	Parietal	Infratentorial	Generalised	Total
Controls	15 (75%)	-	-	-	1 (5%)	-	1 (5%)	17 (85%)
AD	1 (5%)	-	1 (5%)	3 (15%)	3 (15%)	-	3 (15%)	11 (55%)
DLB	5 (25%)	-	-	2 (10%)	2 (10%)	-	6 (30%)	15 (75%)
Tau	-	1 (14%)	2 (29%)	-	1 (14%)	-	-	4 (57%)
TDP43	2 (17%)	1 (8%)	6 (50%)	1 (8%)	-	-	1 (8%)	11 (92%)
FUS	-	-	1 (100%)	-	-	-	-	-
Total	23 (29%)	2 (3%)	10 (13%)	6 (8%)	7 (9%)	-	11 (14%)	59 (74%)

Table 5.5: Algorithm agreement in three or more analysts in a subset of scans (n=80). Agreement was reached in 59 of 80 scans. Complete agreement (among five analysts) was reached in 25 scans (31%). Values are shown as count (group proportion)

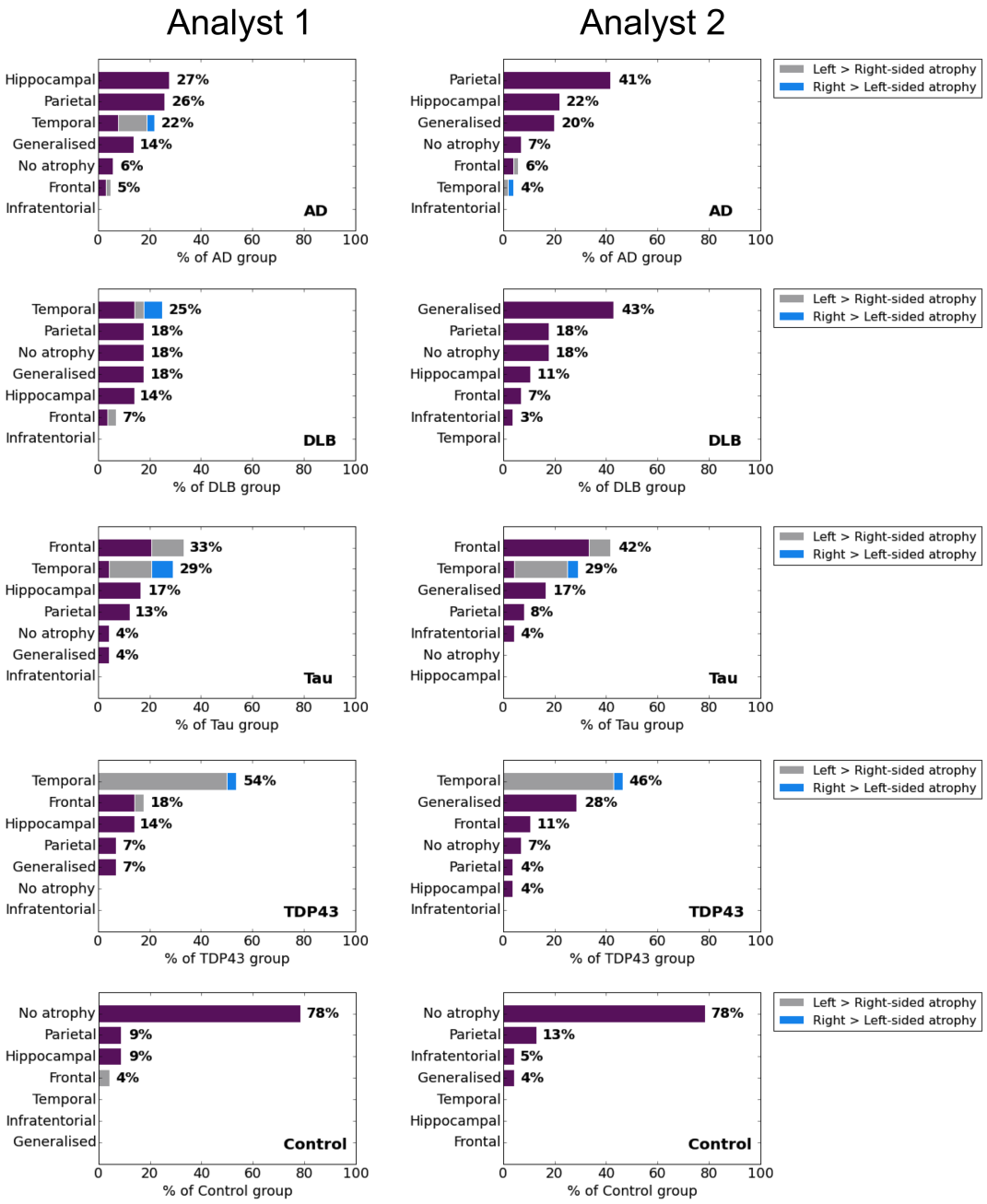


Figure 5.2: Pathology groups contained in the complete dataset (n=207) distributed by most prominent atrophic brain region. Asymmetry within the frontal and temporal branches is displayed as hatched regions.

majority of Alzheimer's disease cases with hippocampal atrophy, while analyst 2 classified almost the same proportion with generalised atrophy. Analyst 3's choices were between the other two, splitting cases evenly between hippocampal and generalised atrophy, followed by parietal lobe atrophy. In the DLB group, analysts 1 and 3 classified the highest proportion of patients with no atrophy and 25% with hippocampal atrophy, while analyst 3 interpreted a more generalised pattern or no atrophy. In the FTLD-Tau cases, analyst 1 detected predominantly frontal and temporal atrophy with a high degree of asymmetry, analysts 2 and 3 categorised most cases as generalised or temporal. In the single FUS case, all analysts agreed on temporal lobe atrophy. All experienced analysts detected a predominantly left-sided asymmetry in the FTLD-TDP43 group and no atrophy in the majority of control cases.

5.4 Discussion

In this chapter, the ability to guide visual image assessment of cerebral atrophy, by synthesising guidance from consensus criteria and findings from the literature in a structured algorithm, was tested against ante mortem imaging acquired in pathologically proven cases. The findings suggest that while characteristic patterns of atrophy associated with each pathology typically predominated at the group level, there is considerable variation at the individual patient level. The consistency among analysts in identifying the brain region exhibiting the most predominant volume loss was moderate at best, with even the experienced analysts only reaching complete agreement in 40% of cases.

Based on the Landis and Koch interpretation of kappa values [287], the confidence intervals calculated for both the complete dataset and the subset of images, consistently straddled the fair to moderate agreement categories. Point estimates of agreement were around κ =0.44 for the two analysts assessing the complete dataset, and both the experts and all analysts assessing the 80-image subset, suggesting a consistent level of variability when applying this technique. The highest measure of agreement κ =0.57 was between two experts who have worked together for a number of years (NF and JS), suggesting that training and experience may improve reliability.

Assessment of frequency of distribution at the individual analyst level suggests that in

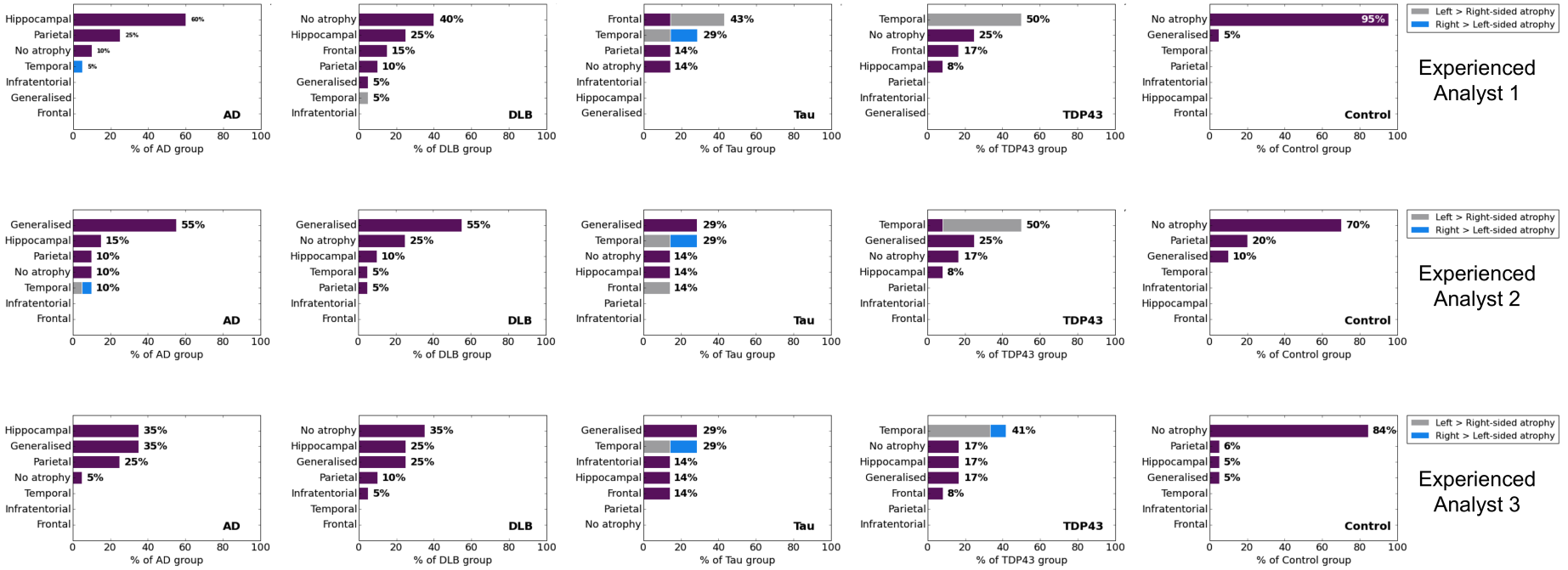


Figure 5.3: Pathology groups contained in the data subset (n=80) distributed by most prominent atrophic brain region. Asymmetry within the frontal and temporal branches is displayed as hatched regions.

most cases, the characteristic patterns of atrophy associated with underlying pathology (as described in the literature) predominate at the group level. The highest proportion of Alzheimer's disease cases were consistently associated with disproportionate hippocampal or parietal lobe atrophy [37, 39, 109, 111, 139], while the FTLD pathologies were associated with (often asymmetric) frontal and temporal lobe atrophy [59, 60]. DLB pathology was less well defined with less hippocampal atrophy than was reported in the Alzheimer's disease group [244, 248–251]. In this context, one possible explanation for the relatively low inter-rater reliability is that signature patterns of atrophy often span lobar regions, and it may not be possible to determine which of these regions is more severely affected based on visual inspection alone. Additionally, distinct regions may be equally affected, such as the hippocampi and parietal lobes in some patients with Alzheimer's disease. Many of the analysts identified a generalised non-specific pattern of atrophy within the images; however, it is worth noting that the most experienced analyst was more definitive in his choices and less reliant on this category. This analyst's choices better reflected the evidence in the literature, suggesting that perhaps this technique is more suitable for experienced users.

While this approach provides a useful summary of the existing evidence for characteristic patterns of cerebral atrophy associated with neurodegenerative dementia, and encourages critical evaluation of imaging in these regions, it has a number of limitations. Primarily, it may not provide enough specific anatomical landmarks for less experienced users, and it may be overly restrictive in forcing a clear decision between the most affected lobar regions. Often atrophy occurs in the boundary regions between lobar division, or equally in distinct brain regions, and it may be difficult to make a definitive choice. Furthermore, less experienced users may find it difficult to even identify these boundaries. More experienced analysts are more likely to incorporate this information into their assessment to reach an appropriate diagnosis, however, this decision making process is difficult to operationalise in a simple algorithm. The algorithm also attempts to separate hippocampal atrophy from temporal lobe atrophy to better discriminate between Alzheimer's disease and FTLD pathologies; however, this may not be practical in later stages of disease as focal atrophy expands to include surrounding brain regions. The primary limitation of this study protocol is the relatively

small number of images assessed by the experienced analysts, and in particular the small sample of FTLD-Tau and FTLD-TDP43 cases included.

5.5 Conclusion

Certain characteristic patterns of cerebral atrophy, with predictive value for the neurodegenerative dementias can be identified from unstructured visual assessment of 3D T1-weighted imaging. However, both skill and experience are required to identify these patterns consistently and even in experienced hands, there is relatively poor agreement in identifying the brain regions most affected by atrophy. A more targeted, quantitative approach, introducing specific anatomical landmarks, may improve diagnostic accuracy and make this information more accessible to a wider clinical audience. In the following chapters a hierarchy of such techniques, from visual rating to fully automated segmentation, are investigated for their ability to fill this requirement.

Chapter 6

Evaluation of Visual Rating Scales

6.1 Introduction

Visual scan assessment remains the primary method for extracting diagnostically useful information in clinical settings. However, as demonstrated in the previous chapter, without operational guidelines to identify, report or interpret patterns of atrophy with diagnostic value in dementia much potentially relevant information may be underutilised. Visual rating scales, specifically designed to assess general and focal cerebral atrophy in patients with cognitive impairment (reviewed in Harper et al [120]), may provide such a framework, allowing for the reliable identification and interpretation of imaging findings of value in the differential diagnosis of dementia. Furthermore, since visual rating scales are both quick and easy to apply, and can be performed on routinely acquired images, they offer an inexpensive means of extracting this information, ideally suited for implementation into clinical practice, and may make it easier for clinicians without neuroradiology expertise to extract diagnostically useful information.

Several visual rating scales have been developed specifically to rate brain regions vulnerable to atrophy in a range of different dementias. While some have been used extensively in both research and in clinical settings, most notably the Scheltens medial temporal lobe scale [129], many have only been evaluated in small single centre studies. Few studies have attempted to directly compare or to combine the diagnostic value of individual scales, fewer still in a large multi-centre setting to determine the real world generalisability and robustness of such findings, and to our knowledge, no study has exclusively assessed their diagnostic utility when applied to scans acquired from individuals with pathologically confirmed dementias.

In this chapter, using structural MR scans from healthy individuals and a large sample of patients with a histopathological diagnosis of dementia, the aim was to: (1) eval-

uate the reliability of six different visual rating scales and the time taken to perform these ratings; (2) explore the relationship between each visual rating scale and pattern of grey matter volume loss; (3) compare the performance of rating scales to expert scan assessment in predicting underlying pathology; and (4) determine if a machine learning (support vector) approach, based on all visual rating scale scores, can improve prediction accuracy.

6.2 Methods

6.2.1 Study Population

The study population is described in Section 4.2. In brief, it consists of 184 patients with a diagnosis of dementia during life and pathological confirmation of underlying pathology: 101 had a primary diagnosis of Alzheimer's disease, 28 with DLB, and 55 with FTLD (24 tauopathies, 28 TDP43 proteinopathies, 3 fusopathoies). 73 cognitively normal control subjects were also included in the analysis.

6.2.2 Structural MRI

Image acquisition is described in Section 4.3. All individuals had 3D T1-weighted volumetric MR imaging acquired during life.

6.2.3 Visual Rating of Cerebral Atrophy

Visual rating of the complete imaging dataset of all patients and controls (n=257) was performed, blind to all clinical and pathological information, by two trained raters (Dr Giorgio Fumagalli and Lorna Harper). Three regions were rated based on existing scales previously described in the literature: the five-point anterior temporal (AT) scale by Kipps/Davies *et al* [125, 126]; the five-point medial temporal lobe atrophy (MTA) scale by Scheltens *et al* [129] currently recommended in the research guidelines for the diagnosis of Alzheimers disease [138]; and the four-point posterior atrophy (PA) scale by Koedam *et al* [139]. To provide additional, more fine-grained assessment of anterior atrophy, an adapted and simplified version of a visual rating scale originally devised by Davies *et al* [145] was used, as described by Fumagalli *et al* [288]. In brief, three

regions orbito-frontal (OF), anterior cingulate (AC) and fronto insula (FI) previously shown to have potential for differential diagnosis [127, 145, 209] were selected. To improve usability, each scale was simplified to a 4-point scale, and reference images were devised. To improve consistency, slice selection was specified, with the OF and AC regions both rated on the first anterior slice where the corpus callosum becomes visible, and the fronto-insula rated over three slices, starting on the first anterior slice where the anterior cingulate becomes visible and moving posteriorly. Images were rated in native space, in keeping with standard clinical reads. To aid rating consistency, reference images for each rating scale were provided to the raters (see Appendix E). Separate scores were recorded for regions in left and right hemispheres.

To provide independent validation of the results from the two primary raters, two visual rating experts (Frederik Barkhof and Philip Scheltens) also assessed 80 scans (20 Alzheimers disease, 20 DLB, 20 FTLD, 20 controls) drawn at random from the total study population. GF also re-rated this subset population. The time taken by each rater to apply the visual rating protocol to each image was automatically recorded to estimate the feasibility of implementing such a protocol in clinical practice.

6.2.4 Expert Diagnosis

Six clinical dementia experts (FB, NF, JO, PS, JS, GF), each provided what they thought was the most likely pathology diagnosis for the above-mentioned subset study population (n=80) based on independent, unstructured visual assessment of each MR image. Experts were blinded to all clinical and pathological information except the persons age at the time of scanning. Images were displayed in a random sequence in terms of underlying pathology.

6.2.5 Voxel Based Morphometry

VBM was used to explore the relationship between each rating scale and its associated pattern of grey matter volume loss. VBM pre-processing is described in detail in Section 2.3.4 and Section 4.4.4. In brief, images were segmented into grey matter, white matter and CSF, spatially normalised using DARTEL, and smoothed (full-width half-maximum 6mm).

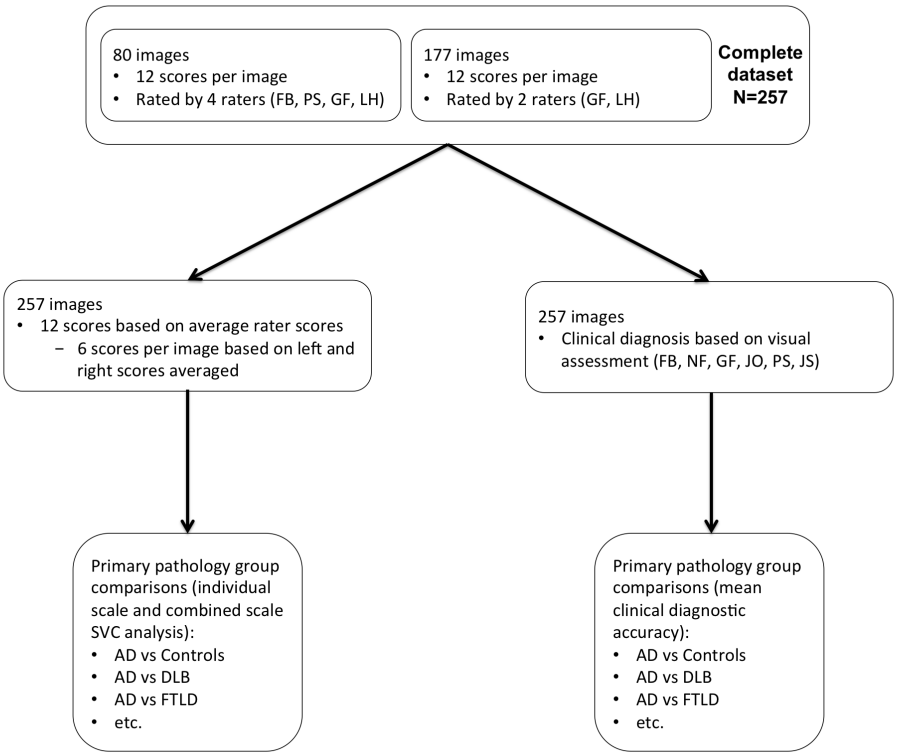


Figure 6.1: Flowchart demonstrating how the visual assessment data (visual ratings and expert diagnosis) was generated and combined for further analysis.

6.2.6 Statistical Analysis

Figure 6.1 illustrates how the visual assessment data (visual ratings and expert diagnosis) was generated and combined for further analysis.

Inter-rater reliability of each rating scale was determined using the intraclass correlation coefficient (ICC). As described by Shrout and Fleiss [289] there are several forms of ICC, with the appropriate form determined by the underlying statistical model and the intended application of the reliability results. In this study a two-way random, absolute, single-measures ICC (ICC(2,1)) was used to estimate the reliability of each scale when applied by a single rater. ICC(2,1) was calculated separately for the subset group (n=80) based on three raters (FB, GF, PS), and the total study population (n=257) based on two raters (GF, LH). Average measures ICCs (ICC(2,k)) were also calculated to estimate the improvement in reliability of each scale when based on average scores from multiple raters.

Partial correlation of grey matter volume with mean visual rating scores (based on the mean scores from four raters (FB, GF, LH, PS) in the subset population and two raters (GF, LH) in the remainder of the total population) was assessed by applying the general linear model at the level of each voxel using all images (n=257). Left and right hemisphere scores were averaged for each scale such that grey matter volume was modelled as a function of the six rating scales (OF, AC, AT, FI, MTA, PA) and adjusted for age, gender, and TIV by including these variables as covariates in the model (equation given in Figure 6.2). Six additional models were also created to investigate simple correlation of each individual scale with grey matter volume, also adjusted for age gender and TIV (equations given in Figure 6.4). A mask was created, based on the optimal threshold of the group average image, using the automatic mask creation strategy in the SPM toolbox [165]. Correction for multiple comparisons was made by using random field theory to control the family-wise error rate at a significance level of 0.05.

Expert rater diagnosis was assessed for each binary disease group comparison and reported in terms of sensitivity, specificity, and balanced accuracy (0.5*(sensitivity+specificity)). The ability of each visual rating scale (averaging left and right scores) to predict pathology was similarly assessed, with the addition of the area under the receiver-operator characteristic curve (AUC). support vector classifier (SVC) accuracy was also determined using balanced accuracy and AUC. Confidence intervals were calculated using the Hanley-McNeil approach evaluated in Newcombe *et al* [290]. Data processing and analysis was performed using Python libraries NumPy 1.8.1, SciPy 0.14.0 and Pandas 0.14.1 on Python 2.7.6 64-bit.

Linear SVC were used to predict pathology diagnosis based on mean visual rating scores calculated for the entire study population. Using mean left and right sided scores from each of the six visual rating scales resulted in 6 features in total. Twelve SVC classifiers were used to assess each binary comparison: e.g. the prediction of Alzheimers disease pathology from the control group; Alzheimers disease from DLB; Alzheimers disease from FTLD; and Alzheimers disease from DLB+FTLD, etc. Features were corrected for age at the time of scanning. Split-half separation was used to divide the data for each classifier into training and testing sets. The training data was scaled to zero mean and unit variance over subjects, with the same transformation then applied

to the testing data. SVCs were trained using leave-one-out cross-validation and class weighting was applied to adjust for unbalanced groups. The regularisation parameter, C, was optimised using grid-search in the range 1E-5 to 100, increasing by an order of magnitude each time. The SVC was implemented using the squared-hinge loss function and L2 regularisation. Classification accuracy is presented as balanced accuracy and AUC values. Feature weighting for each classifier is discussed as an indication of each scales contribution to group separation [291]. SVC processing and analysis was performed using the Python libraries SciPy 0.14.0 and Scikit-Learn 0.15.2 [286] on Python 2.7.6 64-bit.

6.3 Results

6.3.1 Demographics

Demographic details of the patients and control subjects are shown in Table 6.1. The subjects were well matched for gender, disease duration and TIV. The DLB subjects were significantly older ($p<0.001$), with less time between scan until death ($p<0.05$) than the Alzheimers disease and FTLD patients. MMSE within six months of scan date was only available in 116 of the 184 patients (missing data in 37/101 Alzheimers disease, 5/28 DLB, 26/55 FTLD). Based on the data available, MMSE was significantly higher in the FTLD than the Alzheimers disease group ($p<0.001$). Very similar results were found when using MMSE closest to scan (n=170/184).

6.3.2 Expert Diagnoses Based on Unstructured Visual Scan Assessment

The mean sensitivity, specificity and balanced accuracy of the six expert diagnosis based on standard, unstructured assessments of the images are shown in Table 6.2. Balanced accuracy was high (~90%) for distinguishing Alzheimers disease and FTLD from controls, and ~70% for DLB vs controls. For the more clinically relevant head-to-head disease comparisons balanced accuracy was on the order of 70-80%. Balanced accuracy for the multiple disease group comparisons was ~60-70%, with specificities of 69-86%, but sensitivities were more variable, ranging from 34-67%.

	Control	AD	DLB	FTLD
Demographics				
N	73	101	28	55
Gender (%M)[A]	52	61	75	56
Age at scan (yrs)	67 (8)	61 (11)	70 (6)	61 (9)
Disease duration at scan (yrs)	-	4 (3)	3 (2)	3 (3)
Time from scan until death (yrs)	-	6 (3)	4 (2)	5 (3)
MMSE (x/30)	-	18 (6)	20 (5)	23 (6)
TIV (mls)	1501 (159)	1479 (150)	1550 (148)	1498 (149)
Mean visual rating scores				
Orbito-frontal	0.9 (0.5)	1.6 (0.8)	1.5 (0.7)	2.3 (0.8)
Anterior cingulate	1.0 (0.5)	1.3 (0.7)	1.3 (0.5)	1.9 (0.8)
Fronto-insula	1.2 (0.5)	1.7 (0.6)	1.6 (0.5)	2.1 (0.6)
Anterior temporal	0.9 (0.5)	1.5 (0.5)	1.3 (0.4)	2.1 (0.9)
Medial temporal	0.6 (0.6)	1.6 (0.9)	1.1 (0.7)	2.2 (0.9)
Posterior	0.9 (0.7)	1.6 (0.9)	1.2 (0.8)	1.3 (0.6)

Table 6.1: Patient demographics and mean visual rating scores for each group. Data are reported as mean (standard deviation). TIV=Total intracranial volume, MMSE=Mini mental state examination.

Classification Task	Sensitivity	Specificity	Balanced Accuracy
AD from Controls	92% (7%)	86% (10%)	89% (6%)
DLB from Controls	49% (20%)	92% (6%)	70% (9%)
FTLD from Controls	82% (5%)	99% (3%)	90% (2%)
AD from DLB	84% (9%)	58% (17%)	71% (6%)
AD from FTLD	81% (12%)	74% (11%)	77% (10%)
FTLD from DLB	89% (8%)	75% (26%)	82% (12%)
AD from DLB+FTLD	67% (13%)	69% (4%)	68% (8%)
DLB from AD+FTLD	34% (16%)	86% (6%)	60% (7%)
FTLD from AD+DLB	59% (9%)	80% (11%)	69% (6%)

Table 6.2: Accuracy of expert diagnosis based on visual assessment of structural imaging (n=80, 20 per group). Experts were blinded to all clinical and pathological information except the persons age. Data are presented as mean (standard deviation) based on six dementia experts. AD=Alzheimers disease DLB=dementia with Lewy bodies, FTLD=frontotemporal lobar degeneration.

6.3.3 Time to Perform Visual Rating

Mean time to perform and record all six visual rating scales based on three raters assessing the subset study population (n=80) was 2.9$\pm$1.3 minutes. Individual rater means and standard deviations were 2.7$\pm$1.1, 2.4$\pm$1.0, and 3.6$\pm$1.6 minutes.

6.3.4 Inter-rater Reliability of Visual Rating Scores

Single measure and average measure ICC results for each scale are shown in Table 6.3. For the single measures ICC values, representing the reliability of each scale at the level of the individual rater, the MTA scale performed best overall, with very similar results achieved with two raters assessing all 257 scans, and three raters scoring 80 scans (ICC(2,1) $\geq$ 0.79). The PA, OF and FI scales also demonstrated good reliability (ICC(2,1) $\geq$ 0.71) based on two raters assessing the total study population; reliability

was slightly reduced when performed by three raters in the subset population (ICC(2,1) $\geq$ 0.56). Conversely, the reliability of the AT scale based on the subset population (ICC(2,1) $\geq$ 0.72) was greater than the reliability based on the total study population (ICC(2,1) $\geq$ 0.57). The reliability of the AC scale was lowest overall (ICC(2,1) range: 0.56-0.62). As expected, the reliability based on mean rater scores was consistently greater for all scales (ICC(2,k) $\geq$ 0.73). There were no material differences in reliability based on the larger or smaller population samples for any scale with the exception of the AT scale which was less reliable in the larger population sample.

6.3.5 Correlation of Grey Matter Volume with Visual Rating Scores

VBM analysis revealed a negative partial correlation of higher visual rating score with lower grey matter density for all visual rating scales. As shown in Figure 6.2, the pattern of regional atrophy correlated very closely with the specific brain region each scale was designed to assess. This regional specificity was highest for the MTA scale, although even the smaller frontal regions (OF and AC) showed significant correlation with their visual rating scales. Only the AT scale demonstrated a small region in the left superior parietal lobule/supramarginal gyrus where visual rating scores were positively correlated with grey matter volume loss (i.e. the reverse contrast, see Figure 6.3). Since higher AT scores are associated with FTLD pathologies, and in particular TDP43-C pathology associated with semantic dementia [235], which are less likely to demonstrate atrophy in posterior brain regions, this result is pathologically plausible. As expected, analysis of each scale in separate models (without influence of the other scales) demonstrated a more diffuse pattern of atrophy, although, the most highly correlated regions were still confined to, or included, the brain region targeted by each scale (see Figure 6.4).

6.3.6 Pathology Classification Accuracy for Each Visual Rating Scale

Mean visual rating scores for each of the scales were significantly different between pathology groups, primarily driven by higher scores in the FTLD group in all but the

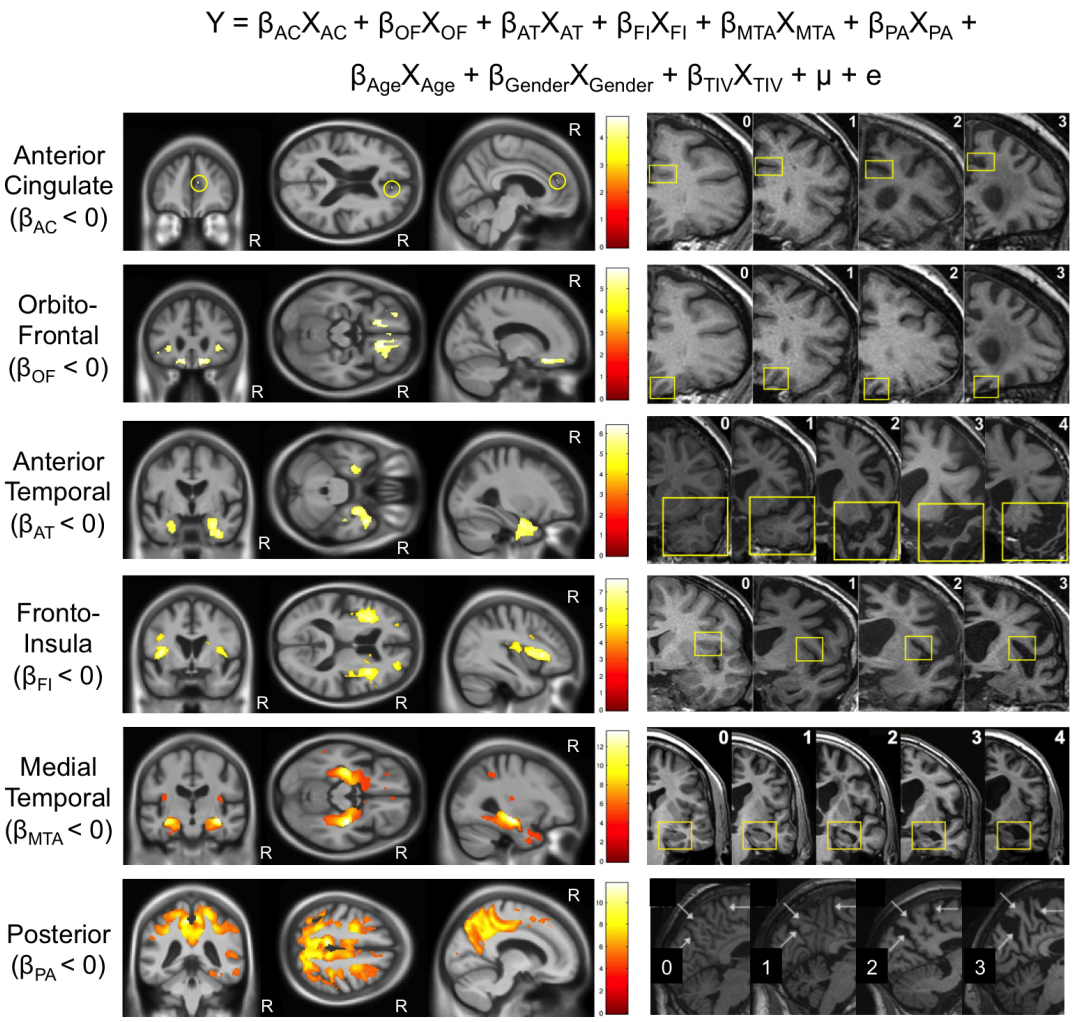


Figure 6.2: Voxel-based morphometry images demonstrating negative partial correlation between grey matter volume and each visual rating scale, adjusted for the other scales (modelled by the equation above). In all images statistical significance of correlations was corrected for multiple comparisons (family wise error rate $p<0.05$). The corresponding visual rating scale reference images are displayed adjacent to each statistical parametric map. R indicates the right hemisphere.

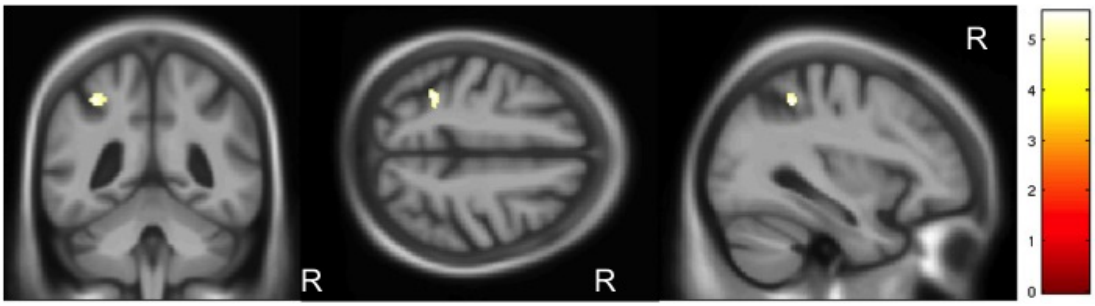


Figure 6.3: VBM analysis of the anterior temporal scale demonstrated a small region in the left superior parietal lobule/supramarginal gyrus where visual rating scores were positively correlated with grey matter atrophy. Statistical significance was corrected for multiple comparisons (family wise error rate p<0.05)

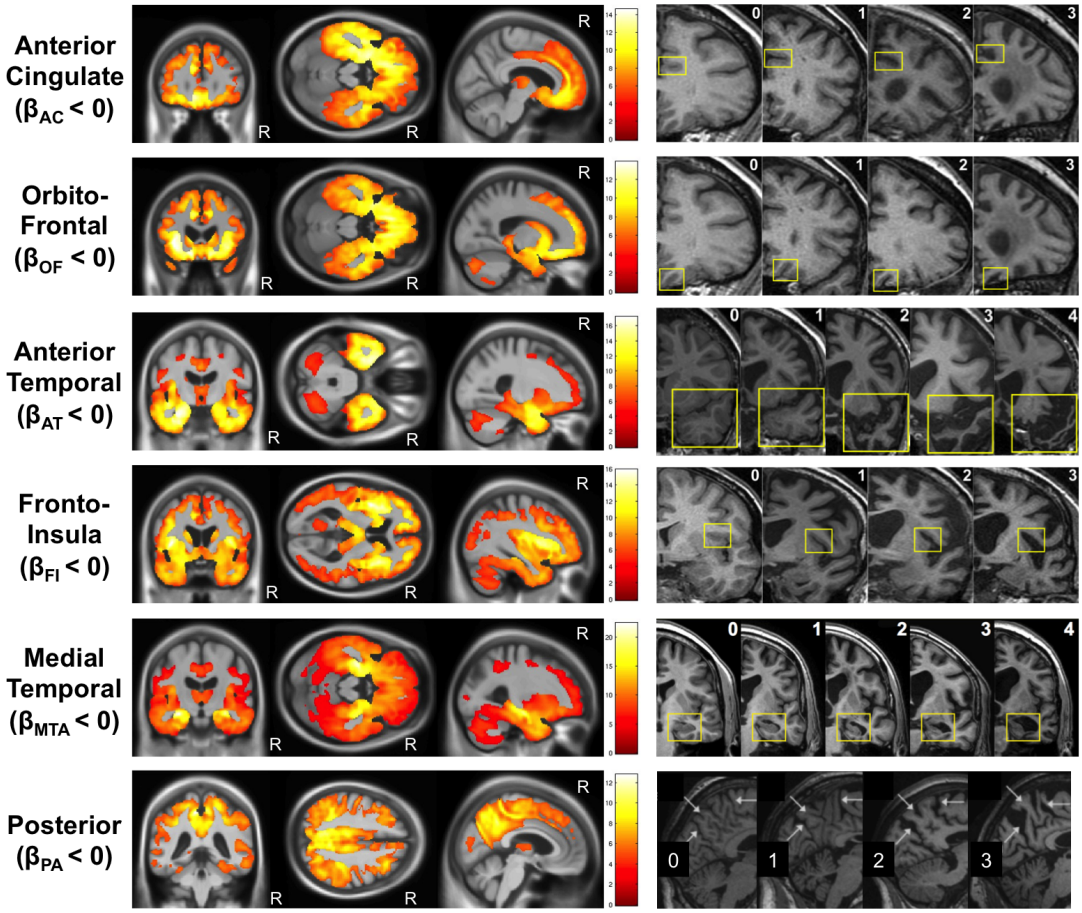


Figure 6.4: Six simple correlations of each scale with grey matter volume, i.e. $y = \beta_{AC}X_{AC} + \beta_{Age}X_{Age} + \beta_{Gender}X_{Gender} + \beta_{TIV}X_{TIV} + \mu + e$, etc. In all images statistical significance of correlations was corrected for multiple comparisons (family wise error rate p<0.05). The corresponding visual rating scale reference images are displayed adjacent to each statistical parametric map. R indicates the right hemisphere.

	Single-Measures ICC		Average-Measures ICC	
Scale	**3 raters, n=80**	**2 raters, n=257**	**3 raters, n=80**	**2 raters, n=257**
LMTA	0.84 (0.77-0.89)	0.84 (0.79-0.87)	0.94 (0.91-0.96)	0.91 (0.89-0.93)
RMTA	0.79 (0.7-0.86)	0.83 (0.79-0.86)	0.92 (0.88-0.95)	0.91 (0.88-0.93)
LPA	0.57 (0.44-0.69)	0.71 (0.65-0.77)	0.80 (0.7-0.87)	0.83 (0.79-0.87)
RPA	0.65 (0.53-0.75)	0.72 (0.66-0.78)	0.85 (0.77-0.9)	0.84 (0.79-0.87)
LAT	0.78 (0.69-0.85)	0.62 (0.53-0.70)	0.91 (0.87-0.94)	0.77 (0.70-0.82)
RAT	0.72 (0.61-0.8)	0.57 (0.48-0.65)	0.88 (0.83-0.92)	0.73 (0.65-0.79)
LOF	0.6 (0.47-0.71)	0.72 (0.65-0.78)	0.82 (0.73-0.88)	0.84 (0.79-0.88)
ROF	0.62 (0.50-0.73)	0.74 (0.68-0.79)	0.83 (0.75-0.89)	0.85 (0.81-0.88)
LAC	0.56 (0.43-0.68)	0.61 (0.51-0.69)	0.79 (0.7-0.87)	0.76 (0.68-0.82)
RAC	0.57 (0.44-0.69)	0.62 (0.53-0.69)	0.80 (0.7-0.87)	0.76 (0.69-0.82)
LFI	0.56 (0.42-0.68)	0.73 (0.67-0.79)	0.79 (0.69-0.87)	0.85 (0.80-0.88)
RFI	0.58 (0.46-0.7)	0.72 (0.65-0.77)	0.81 (0.71-0.87)	0.84 (0.79-0.87)

Table 6.3: Intraclass correlation coefficient (ICC) measures with confidence intervals. Single-measures ICC provide an estimate of the reliability of the scale when applied by a single rater. The average-measures ICC is an estimate of the reliability of the scale based on an average score derived from k raters. **Abbreviations:** L = left, R = right, MTA = medial temporal lobe atrophy, PA = posterior atrophy, AT = anterior temporal, OF = orbitofrontal, AC = anterior cingulate, FI = fronto insula.

PA scale (see Table 6.1). The results for the best performing scale (based on AUC score) for each group comparison are summarised in Table 6.4. The MTA scale was most effective at accurately identifying Alzheimers disease pathology from the control group (AUC=0.82) and the DLB group (AUC=0.67). Higher PA scale scores ($\geq$2.5) added some value in comparisons with the FTLD group, although sensitivity was low (22%). The OF scale was useful for distinguishing DLB from the control group (AUC=0.74). All other scales were below chance at detecting DLB from the other disease groups. The MTA scale was the most effective at identifying FTLD pathology when compared with the control group (AUC=0.92) and the DLB group (AUC=0.81). Higher OF scale scores ($\geq$2.5) were specific for FTLD pathology (81%) when compared with the Alzheimers disease group (AUC=0.73).

6.3.7 Support Vector Classification Accuracy for Pathology Based on Visual Rating Scores

The results for each group comparison using the mean right/left scores (i.e. six features) are summarised in Table 6.4B and illustrated in Figure 6.5 and Figure 6.6. For all comparisons, the balanced accuracy ranged from 61-93%, and AUC 0.67-0.97. SVC classification accuracy demonstrated a substantial improvement over the best single score in all cases, equivalent to or better than expert diagnosis. Based on the feature weighting applied by each SVC, the MTA, PA and AT scales contributed most to the separation of the Alzheimers disease group from controls, and the OF and AT scales contributed most to the separation of the DLB group from controls. With the exception of the PA scale, most scales contributed equally to the separation of the FTLD group from the controls. The PA (indicating Alzheimers disease), AT and OF scales (indicating FTLD) contributed most to the separation of the Alzheimers disease and FTLD groups. All scales except the PA scale contributed similarly to the separation of DLB and FTLD, weighted towards the FTLD group.

6.4 Discussion

This large, multi-centre study of pathologically proven dementias, demonstrates that visual rating scales from routinely acquired structural MR images can be reliably rated,

	A. Best Single Visual Rating Scale					B. SVC Performance Based on All Scales			
Classification Task	**Scale (Cut-off)**	**Sensitivity**	**Specificity**	**Balanced Accuracy**	**AUC**	**Sensitivity**	**Specificity**	**Balanced Accuracy**	**AUC**
AD from Controls	MTA (1.5)	64% (56-72%)	89% (83-93%)	77% (69-83%)	0.82 (0.74-0.87)	94% (86-97%)	89% (80-94%)	92% (83-96%)	0.95 (0.87-0.98)
AD from DLB	MTA (1.5)	64% (52-75%)	68% (56-78%)	66% (54-76%)	0.67 (0.55-0.76)	82% (66-91%)	64% (47-78%)	73% (56-85%)	0.75 (0.58-0.86)
AD from FTLD	PA (2.5)	22% (15-30%)	98% (94-99%)	60% (50-69%)	0.60 (0.51-0.69)	88% (77-94%)	56% (42-68%)	72% (59-82%)	0.78 (0.65-0.86)
AD from DLB+FTLD	PA (2.5)	22% (17-28%)	86% (81-90%)	54% (47-61%)	0.55 (0.48-0.62)	90% (81-95%)	32% (22-44%)	61% (49-72%)	0.67 (0.55-0.77)
DLB from Controls	OF (1.5)	57% (45-69%)	84% (73-90%)	70% (58-80%)	0.74 (0.62-0.83)	64% (46-79%)	92% (77-97%)	78% (60-89%)	0.86 (0.69-0.94)
DLB from AD	OF (1.5)	57% (45-68%)	48% (36-59%)	52% (41-64%)	0.49 (0.37-0.60)	86% (70-93%)	66% (49-79%)	76% (59-87%)	0.75 (0.58-0.86)
DLB from FTLD	PA (3.0)	7% (3-16%)	100% (95-100%)	54% (41-66%)	0.46 (0.33-0.59)	93% (78-98%)	89% (72-96%)	91% (75-97%)	0.92 (0.76-0.97)
DLB from AD+FTLD	AC (1.0)	93% (85-97%)	11% (6-19%)	52% (41-63%)	0.43 (0.32-0.54)	79% (63-88%)	72% (55-83%)	75% (59-86%)	0.80 (0.64-0.89)
FTLD from Controls	MTA (1.5)	82% (73-88%)	89% (82-94%)	85% (77-91%)	0.92 (0.85-0.96)	89% (77-95%)	97% (89-99%)	93% (83-97%)	0.97 (0.88-0.99)
FTLD from AD	OF (2.5)	55% (45-64%)	81% (73-87%)	68% (59-76%)	0.73 (0.64-0.80)	81% (69-89%)	67% (53-78%)	74% (61-84%)	0.78 (0.65-0.86)
FTLD from DLB	MTA (2.0)	69% (56-79%)	82% (70-90%)	76% (63-85%)	0.81 (0.69-0.89)	89% (72-96%)	93% (78-98%)	91% (75-97%)	0.92 (0.76-0.97)
FTLD from AD+DLB	AT (2.0)	64% (55-71%)	63% (54-71%)	63% (55-71%)	0.67 (0.58-0.74)	85% (74-92%)	62% (49-73%)	73% (61-83%)	0.78 (0.66-0.87)

Table 6.4: **A.** Performance of visual rating scale that most accurately predicts pathology for each binary group comparison. The optimal cutoff points should be interpreted as: $<$ cutoff=normal, $\geq$ cutoff=abnormal. Sensitivity and specificity values are selected based on the maximum balanced accuracy score. **B.** Support vector classifier (SVC) performance based on mean left/right scores for each of the six visual rating scales. All values in parts A and B are presented with 95% confidence intervals in brackets. AD=Alzheimers disease DLB=dementia with Lewy bodies, FTLD=frontotemporal lobar degeneration.

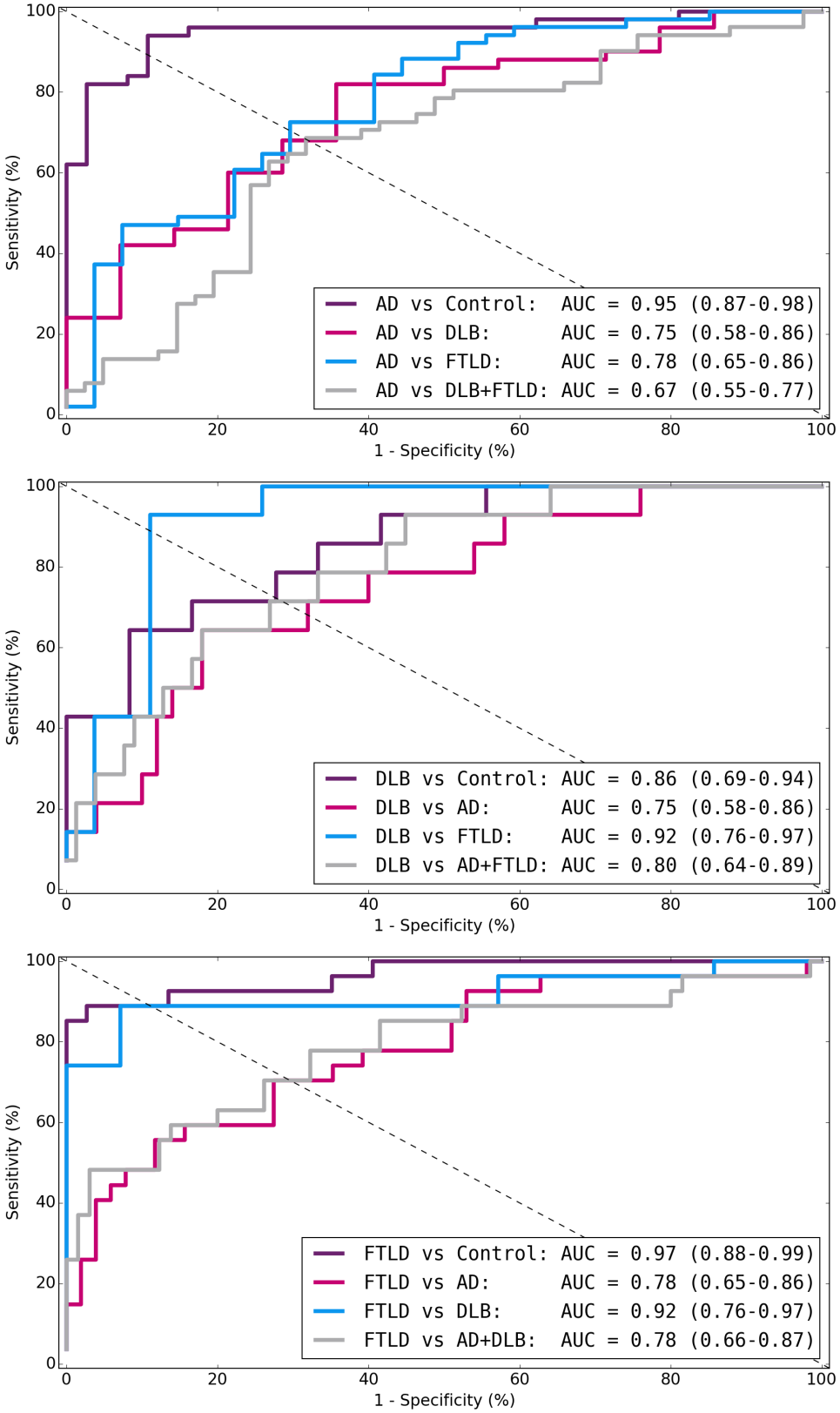


Figure 6.5: Receiver operator characteristic (ROC) plots of support vector classifier (SVC) performance for prediction of **A.** Alzheimers disease (AD), **B.** dementia with Lewy bodies (DLB) and **C.** frontotemporal lobar degeneration (FTLD) pathologies. Area under the curve (AUC) values with 95% confidence intervals are displayed for each classifier.

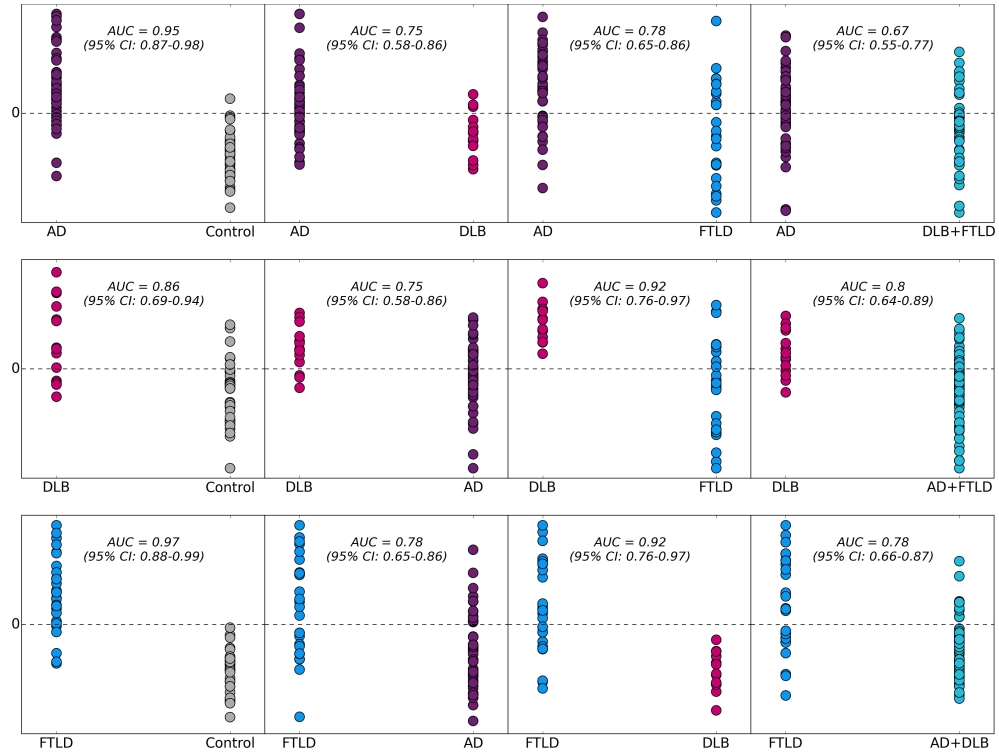


Figure 6.6: Distance plots of support vector classifier (SVC) performance for prediction of dementia pathologies. The points are plotted based on their distance from the separating hyperplane (represented by the dotted line y=0). Area under the curve (AUC) values with 95% confidence intervals are displayed for each classifier.

and are highly correlated with cerebral atrophy in brain regions vulnerable to dementia pathology. Although their contribution to differential diagnosis is somewhat limited when used in isolation, when the scores from all scales are combined in an automated support vector classifier it is possible to achieve diagnostic accuracy equivalent to, and in some cases better than, unstructured scan evaluation performed by expert raters. The rating scales in question are quick and easy to learn and can be applied, in total, in less than three minutes. Taken together, these results suggests that visual rating scales offer clinicians without neuroradiology expertise a means of extracting diagnostically useful information in a time-efficient and inexpensive way, that is ideally suited for integration into routine clinical practice.

The first aim of this study was to directly compare the inter-rater reliability of visual rating scales designed to assess cerebral atrophy in regions particularly vulnerable to the effects of dementia pathology. Although reliability is typically reported in the original concept study of each visual rating scale, and occasionally in follow-up studies, a lack of standardisation in the use and reporting of statistical techniques employed to calculate this metric make it difficult to make direct comparisons (reviewed in Harper *et al* [120]). In this study reliability was investigated using a dataset that is larger and more representative (through the inclusion of multiple dementia pathologies and real life scans acquired on multiple different scanners over many years) than is typically used for this purpose. The MTA scale was consistently highly reliable under all conditions. Of the adapted frontal scales, reliability was higher between the two raters assessing images from the total study population, than in the smaller sample rated by three. This difference is likely to reflect both the differences in sample sizes, and that the two raters had more experience with these scales suggesting that training may improve reliability. The reliability of the AC scale was lower overall, perhaps reflecting the sulcal variability in this rostral region, which can make it difficult to consistently identify the specified region of interest. Of all the scales, the AT scale shows the least linear distribution of atrophy by scale increment, instead adopting a more step-wise change from score 1 to 2 and from score 2 to 3. This lack of differentiation between points 0-1 and 3-4 may account for the variability in reliability between rating conditions. The PA scale, requiring the integration of visual information in three planes in four brain

regions (parietal lobe, posterior cingulate sulcus, parieto-occipital sulcus, precuneus), is undoubtedly the most difficult to apply. Despite this, the two raters assessing the total study population achieved a relatively high degree of reliability, although there was more variability among the three raters based on the smaller dataset. The consistently high reliability of the average measures ICC (based on mean scale scores averaged over raters) perhaps suggests that where possible, the use of mean scores from two or more raters may be preferable, when practicable. However, raters should also undergo training to improve reliability and reduce personal biases. Comprehensive documented instructions may also help to reduce any additional bias from applying multiple rating scales simultaneously.

Using voxel based morphometry, each of the scales was found to be remarkably well correlated with the anatomical regions of interest they were designed to assess, illustrating their regional specificity. This was particularly true for the MTA scale, which was highly associated with hippocampal volume loss, but even the more complex PA scale was well correlated with the posterior pattern of atrophy it was designed to detect. Focal atrophy in the small frontal regions assessed by the OF scale, and to a lesser extent the AC scale (right side only), was also significantly correlated with their associated visual rating scores. Whilst previous studies have investigated the relationship between rating scales and brain volumes in the region of interest [145, 292], these results using an unbiased technique provide independent validation that each of the scales is indeed performing as predicted. The concordance can also be considered as evidence that voxel-based morphometry (as implemented in SPM12b) is performing well in this challenging, heterogeneous dataset.

While several studies have estimated the classification accuracy of rating scales in the diagnosis of various dementias [111, 126, 127, 129, 136, 139, 145, 209, 250, 292] (reviewed in Harper *et al* [120]), very few have used histopathological diagnosis as the gold standard [111, 136, 250], and to our knowledge no study has performed this analysis in such a large, clinically realistic cohort comprising as wide a range of diverse pathologies, and made comparisons with expert scan assessment. There were significant differences in scores between the three canonical pathologies for all visual rating scales. Post-hoc tests revealed significant differences between the control group

and each of the pathology groups ($p<0.05$ with the exception of the AC and PA scale for DLB), reflecting the vulnerability in these regions to dementia pathologies, independent of the affects of ageing. Significant differences in the adapted frontal scales between the disease groups were driven by the FTLD group ($p<0.001$), while the MTA scale demonstrated independent significant differences between each pathology group ($p<0.001$).

As previous studies have shown, the MTA and PA scales were the most useful for predicting Alzheimers disease pathology [111, 129, 139, 250]. We report a lower MTA sensitivity than has been reported in previous studies [129, 250], perhaps due to the relatively young age of this cohort, and the typically higher proportion of non-amnestic, atypical presentations that may be present in younger onset Alzheimers disease [67]. Slightly higher PA score cut-off points ($\geq$2.5), than those reported in previous studies ($\geq$2 [111, 139]), provided optimal separation of Alzheimers disease from FTLD, however, sensitivity for Alzheimers disease pathology (22%) was compromised at the expense of higher specificity (98%), and overall accuracy was slightly lower than previously reported (AUC=0.60 versus 0.66 in [111]. The MTA scale achieved the highest levels of accuracy when distinguishing FTLD pathology from either the control group (AUC=0.92) or the DLB pathology group (AUC=0.81). In agreement with previous work by Hornberger *et al* in a much smaller cohort [209], higher OF scale scores ($\geq$2.5) were best for predicting FTLD from Alzheimers disease (AUC=0.73). The AT scale was the most useful for identifying FTLD pathologies from a pooled group of other dementia pathologies (Alzheimers disease + DLB). The OF scale was reasonably accurate (AUC=0.74) at distinguishing DLB pathology from the control group images. To our knowledge this is a novel application of this scale not previously explored in other studies, but echoes findings in earlier work [293], which found marked and disproportionate frontal atrophy on CT images from autopsy confirmed DLB.

Whilst these results suggest that no single atrophy rating scale can accurately distinguish the common forms of dementia from one another, by combining all scores in an automated classifier it is possible to achieve diagnostic accuracy equivalent to, and in some cases better than, unstructured scan evaluation performed by dementia experts. Furthermore, accuracy of classification based on visual rating is also consistent with

the reported accuracies from a previous study using grey matter volume to distinguish between pathologically confirmed Alzheimers disease and FTLD cases [222]. Given the ease and accuracy of applying these ratings this approach provides a potentially valuable way for non experts to extract valuable diagnostic information from routine scans.

Whilst there is considerable interest in using molecular biomarker techniques to aid in the differential diagnosis, particularly of Alzheimers disease [294], it is notable that the classification accuracy report is comparable to the accuracy of the CSF A1-42 level as recently reported in a large sample by Ewers *et al* (balanced accuracy: Alzheimers disease from DLB = 64%, Alzheimers disease from FTLD = 81%) [295]. Although these tests are identifying different aspects of the disease process, and noting that only the primary post-mortem diagnosis excluding co-pathology was used, each tests contribution to an accurate differential diagnosis is similar. Continued optimisation of the classifier through the inclusion of more data is likely to improve performance beyond what can be achieved with simple dichotomisation of an individual scale. Recent work by Ferreira *et al* [137], and previously by Barkhof *et al* [136] to include age specific cut-offs into visual rating scales could more easily be incorporated into an automated classifier (see also Coupé *et al* [296]; Koikkalainen *et al* [297]) without the requirement to update manual protocols and retrain raters. While the required level of data to achieve such classifiers is unlikely to be available within any single centre, pooling imaging and pathology data between centres and making them accessible online to predict pathology from rating scores could provide a communal resource that is useful for both research purposes and as a diagnostic aid in clinic.

This study has a number of strengths including the large overall sample size, use of multiple scales, post-mortem confirmation of diagnosis, real life acquisition of scans, and comparisons based on blind assessment, i.e. without the benefit of clinical information which in practice is likely to improve diagnostic performance. Limitations include the imbalance in the pathology groups, in particularly the relatively low number of DLB cases included. Furthermore, disproportionate representation of young-onset Alzheimers disease cases in this sample may make the results less generalizable to the average clinical population. To obtain sample sizes sufficient for these analyses, FTLD

was treated as a single diagnostic group rather than as a collection of different pathologies. Control subjects were not pathologically confirmed, therefore, pre-symptomatic pathology cannot be ruled out in this group, which would result in an underestimation of specificity. However, this does not affect the more clinically relevant between-pathology group comparisons. Whilst the sample size is very large in the context of pathologically confirmed dementias, larger numbers in all groups would improve statistical certainty, particularly in the SVC experiments where it is necessary to split the data into training and testing sets. Greater power could also allow for more fine-grained analysis of subtypes of the canonical dementias, and for an investigation of the role of mixed or multiple pathologies. In terms of expert scan assessment, classification performance was based on a subset of the total study population and assumed to represent assessment of the entire dataset. Visual assessment was also performed in native space to better reflect clinical practice, however, reorientation to standard space would allow for greater anatomical consistency between scans and may potentially improve inter-rater reliability and diagnostic accuracy. Variation in image slice thickness could also prevent optimal slice selection for visual rating, however, the use of 3D rather than 2D images, as is used in this study (with typical voxels sizes of less than 2mm) should avoid such issues. Finally, it will be of considerable interest to see if similar results can be obtained by training individuals without prior experience of scan rating to apply these rating scales.

6.5 Conclusions

In summary, this study demonstrates the utility of visual rating scales to provide diagnostically useful information, which when considered in the context of a detailed clinical examination may help to improve the accuracy of clinical diagnosis for the degenerative dementias. Visual rating offers a simple and reliable framework to capitalise on the structural imaging already acquired in most patients at no extra cost. Until more advanced image analysis techniques are adapted for use in clinical practice, the incorporation of visual rating scales (certainly when combined with an automated classifier) offers a quick, simple, reliable means of extracting valuable diagnostic information from structural brain imaging.

Chapter 7

Distinct Patterns of Atrophy in Pathologically Confirmed Dementias

7.1 Introduction

Voxel based morphometry (described in Section 2.3.4) can be used to provide a comprehensive assessment of differences in grey matter volume between predefined groups. It has been used extensively in the study of neurodegenerative dementia, however, the vast majority of studies have explored differences between clinical syndromes, or investigated correlation between grey matter volume and cognition. Very few studies have looked for distinctive patterns of atrophy in pathologically proven cases.

Figure 7.1 summarises the published VBM studies to date based on imaging acquired in pathologically or genetically confirmed dementias. Only 35 studies were found in total after an extensive search of the PubMed database. While around half of these studies provide some assessment of differences in grey matter atrophy between pathologically confirmed dementias and healthy control subjects, this was rarely the focus of the study. Instead, groups are almost always further classified by clinical syndrome to look for distinct imaging signatures associated with cognitive profiles (e.g. aphasia, corticobasal syndrome etc.) [214, 216, 218, 220, 232, 245, 265], and post-hoc analysis of these findings in relation to molecular pathology may not be representative of the group as a whole. Of the studies that were designed to investigate specific patterns of atrophy associated with molecular pathology, more than half are based on group sizes of $\leq$10, and all but two with $<$ 15 [229, 298]. Even fewer studies have been carried out to look directly for differential patterns of atrophy between molecular pathology groups [221, 222, 229, 247, 298], although some additional information can be inferred from studies based on genetic mutation carriers [110, 254, 299].

Detecting consistent patterns of atrophy in larger samples of pathologically proven dementias may provide opportunities to identify regional atrophy not previously considered, and improve diagnostic accuracy at the individual patient level. Regional differences could be used to guide visual image assessment or inform more advanced region of interest analysis using manual or automated segmentation tools. While overlapping patterns of atrophy between pathologies is likely, additional or more severe involvement of particular regions may increase the probability of one pathology over another, and could be used for earlier detection and tracking of disease.

The aim of this chapter, using imaging data from a large sample of pathologically proven dementias, was to investigate consistent patterns of atrophy that differentiate the primary neurodegenerative pathology groups (Alzheimer's disease, DLB, FTLD-tauopathies, FTLD-TDP43 proteinopathies), from healthy control subjects, and critically, from one another. At the time of writing, this is thought to be the first study to include as many molecular pathology groups in the analysis, with larger group sizes than has previously been available, using the most recent and advanced SPM software. A single model is also used to describe the entire dataset in an attempt to reproduce the complexity of the diagnostic challenge often faced in clinic.

7.2 Methods

7.2.1 Study Population

The analysis in this chapter included 181 patients with a diagnosis of dementia during life and post-mortem confirmation of underlying pathology, and 73 healthy control subjects previously described in Chapter 4. Three patients diagnosed with FUSopathies were excluded from the analysis as there was insufficient data to investigate this pathology group.

To allow for more fine grained analysis of dementia subtypes the AD group was subdivided into early-onset (73 early-onset Alzheimer's disease (EOAD)) (age at onset <65 years) and late-onset (28 Late-onset Alzheimer's disease (LOAD)) cases. The population also consisted of 28 DLB cases, 24 FTLD-tauopathy cases and 28 FTLD-TDP43 proteinopathy cases.

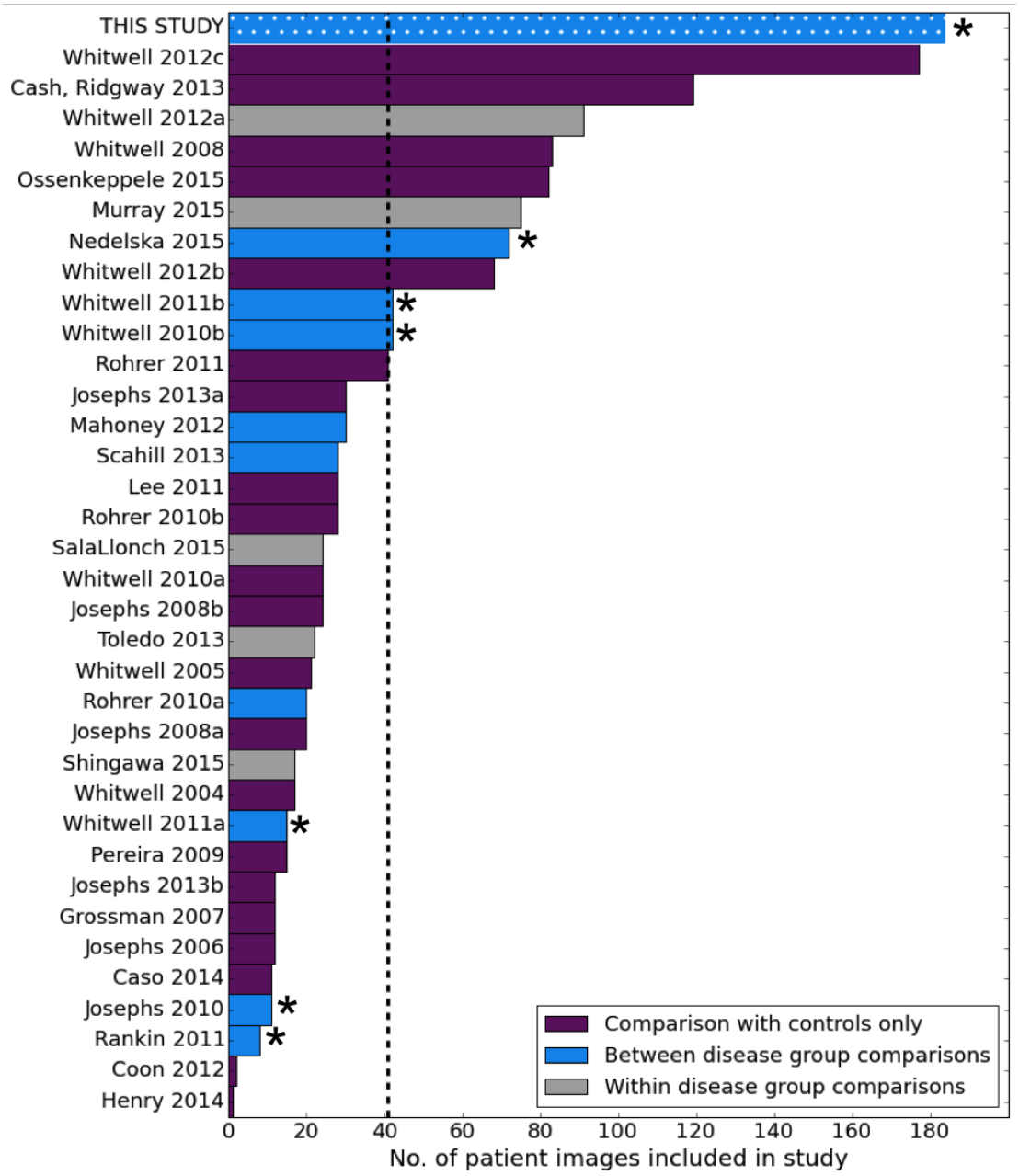


Figure 7.1: Sample sizes of VBM studies using pathologically or genetically confirmed cases. The graph represents the combined total for all pathology confirmed cases included in the study. * indicates between disease comparisons with pathologically (as opposed to genetically confirmed) cases. Mean number of cases across all studies is shown by the dotted line. The study described in this chapter is represented by the dotted bar.

7.2.2 Structural MRI

Image acquisition is described in Section 4.3. All individuals had T1-weighted volumetric MR imaging performed during life.

7.2.3 Voxel Based Morphometry

VBM pre-processing is described in detail in Section 2.3.4 and Section 4.4.4. In brief, images were segmented into grey matter, white matter and CSF, spatially normalised using DARTEL, and smoothed (full-width half-maximum 6mm).

7.2.4 Statistical Analysis

Regional differences in grey matter volume between groups was assessed by applying the general linear model at the level of each voxel using all images (n=254). Grey matter volume was modelled as a function of group and adjusted for age, gender and total intracranial volume by including these variables as covariates in the model. As a retrospective, multicentre study it was not possible to define variables to account for the full extent of scanner and acquisition parameter differences, however, the main source of variability in image quality is likely to be magnetic field strength. Covariates were therefore, also included to account for these differences. Group differences were calculated using one-tailed t-tests (in both directions) between group parameter estimates for each primary group and sub-group comparison.

As described in Section 4.5.2, a mask was created based on the optimal threshold of the group average image, using the automatic mask creation strategy in the SPM toolbox [165]. Unless otherwise stated, correction for multiple comparisons was made using random field theory to control the family-wise error rate at a significance level of 0.001.

In addition to the thresholded statistical parametric maps, differences in grey matter volume between groups that did not reach statistical significance are presented on effect size maps to provide more information about the patterns of cerebral atrophy associated with each pathology group. Effect size maps for each comparison were created by calculating the difference between the estimated beta parameters for each group,

normalised to the control group.

7.3 Results

7.3.1 Demographics

Demographic details of the patients and control subjects are shown in Table 7.1. There were no significant differences between the groups in terms of gender, disease duration and total intracranial volume. The EOAD and TDP43 groups were significantly younger than the other groups ($p<0.05$). MMSE within six months of scan date was only available in 116 of the 184 patients (missing data in 37/101 Alzheimers disease, 5/28 DLB, 26/55 FTLD). Based on the data available, the EOAD had the lowest mean MMSE, significantly lower than the tau group ($p<0.05$). Very similar results were found when using MMSE closest to scan (n=170/184).

7.3.2 Comparison of Disease Groups with Healthy Controls

The effect size maps highlight the fact that all neurodegenerative dementia pathologies cause global loss of brain tissue, with a reduction in volume demonstrated in almost all brain regions when compared with healthy control subjects. As demonstrated in Figure 7.2 and Figure 7.3, these effects extend well beyond those regions that reach statistical significance.

In the SPM images (see Figure 7.2), the Alzheimer's disease group taken as a whole (n=101) demonstrated significant (FWE $p<0.001$) and extensive volume loss compared to the healthy control group, predominantly affecting temporoparietal regions, with regions also affected in the frontal and occipital lobes, and the striatum. These changes were mostly driven by the larger EOAD group (n=73), which demonstrated an almost identical pattern of loss (FWE $p<0.001$). Conversely, the LOAD group (n=28) demonstrated a much more focal atrophy pattern in the medial temporal lobes, centred around the hippocampus (FWE $p<0.001$) (see Figure 7.2).

Although the effect size map demonstrated a broadly similar pattern of volume loss in DLB to the Alzheimer's disease (see Figure 7.3), these effects were less pronounced and are shown on a reduced scale. The statistical maps also demonstrated a much more

	NC	EOAD	LOAD	DLB	Tau	TDP43
N	73	73	28	28	24	28
Gender (%M)	52	59	68	75	58	50
Age (yrs)	67 (8)	56 (8)	75 (6)	70 (6)	64 (9)	60 (8)
Disease duration (yrs)	-	4 (3)	3 (3)	3 (2)	4 (2)	3 (3)
Date of death (yrs)	-	5 (3)	6 (3)	4 (2)	5 (3)	6 (3)
MMSE (x/30)	-	17 (6)	19 (5)	20 (5)	23 (5)	22 (7)
TIV (mls)	1501 (159)	1478 (158)	1482 (132)	1550 (148)	1500 (149)	1474 (151)

Table 7.1: Study population demographics. Results are shown as mean (standard deviation). MMSE is within six months of imaging acquisition

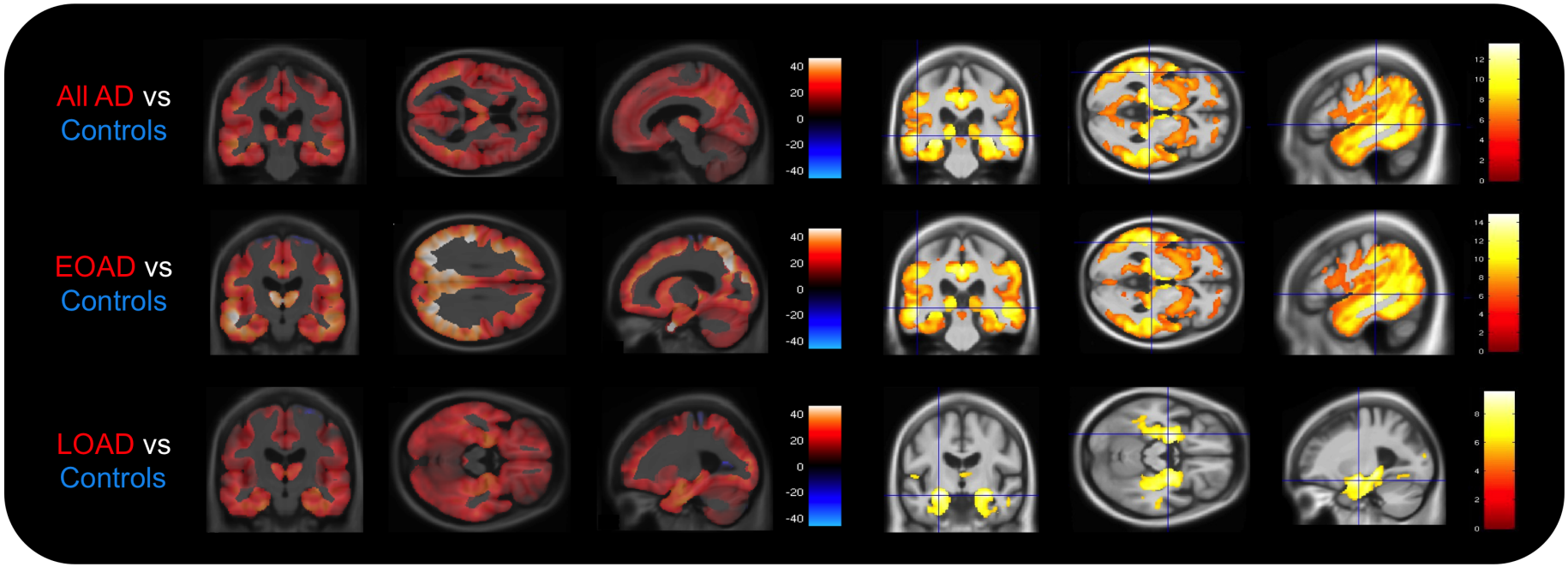


Figure 7.2: Effect size maps are shown on the left, with statistical parametric maps on the right. SPMs are corrected for multiple comparisons using random field theory to control the family-wise error rate at a significance level of 0.001. The crosshairs represent the global maximum difference. The top row represents the AD group taken as a whole (EOAD+LOAD).

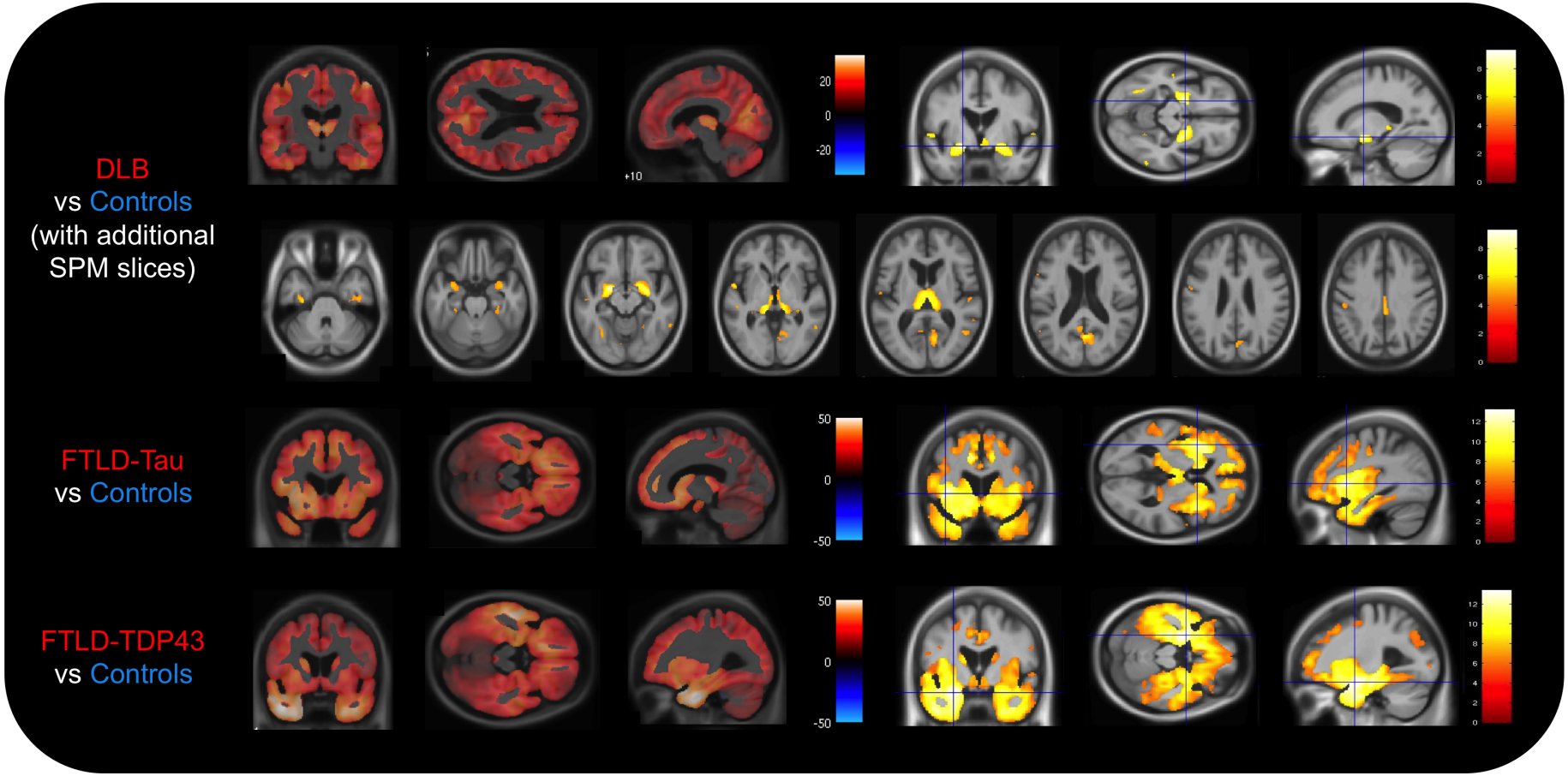


Figure 7.3: Effect size maps are shown on the left, with statistical parametric maps on the right. Additional SPM slices are presented for the more subtle DLB vs control group comparison. SPMs are corrected for multiple comparisons using random field theory to control the family-wise error rate at a significance level of 0.001. The crosshairs represent the global maximum difference.

subtle pattern of grey matter volume loss than the Alzheimer's disease group, affecting the thalamus, the amygdala, and regions around the fusiform gyrus and posterior cingulate sulcus, with relative sparing of the hippocampus (see Figure 7.3).

The FTLD pathology groups demonstrated the most marked differences in brain volume, and the respective effect size maps are shown on a larger scale than the Alzheimer's disease and DLB groups (see Figure 7.3). In both the FTLD-Tau and FTLD-TDP43 groups, significant volume loss was demonstrated in frontal and temporal lobe regions (FWE $p<0.001$), however, volume loss in the FTLD-Tau group was more marked in anterior and superior frontal lobe regions, with extension into the anterior temporal lobes, while there was greater extension into the temporal lobes in the FTLD-TDP43 group. Both groups demonstrated subtle asymmetry with the left hemisphere slightly more affected than the right.

7.3.3 Comparison of AD with DLB

There were very few significant differences detected between the Alzheimer's disease and the DLB groups (see Figure 7.4). The Alzheimer's disease group demonstrated greater tissue loss in some very small regions in the parietal lobe and around the posterior cingulate that survived multiple comparisons correction ($p<0.05$ FWE). The DLB group showed greater volume loss in small regions around the pre- and post-central gyri, however, they did not survive correction for multiple comparisons. The effect size map also highlights the pre- and post-central gyri as being more affected in DLB, with some additional regions in the occipital lobe and the cerebellum. All other brain regions were more severely affected in patients with DLB pathology.

7.3.4 Comparison of AD with FTLD Pathologies

The FTLD-Tau group had significantly less grey matter volume in frontal lobe regions than the Alzheimer's disease group ($p<0.001$ FWE) (see Figure 7.6). In particular, grey matter volume was reduced in orbitofrontal regions, the anterior cingulate and the insula, with the left hemisphere slightly more affected than the right. The reverse contrast only demonstrated a very small region in the right parietal lobe where volume loss was greater in the Alzheimer's disease group than the FTLD group ($p<0.05$ FWE). The

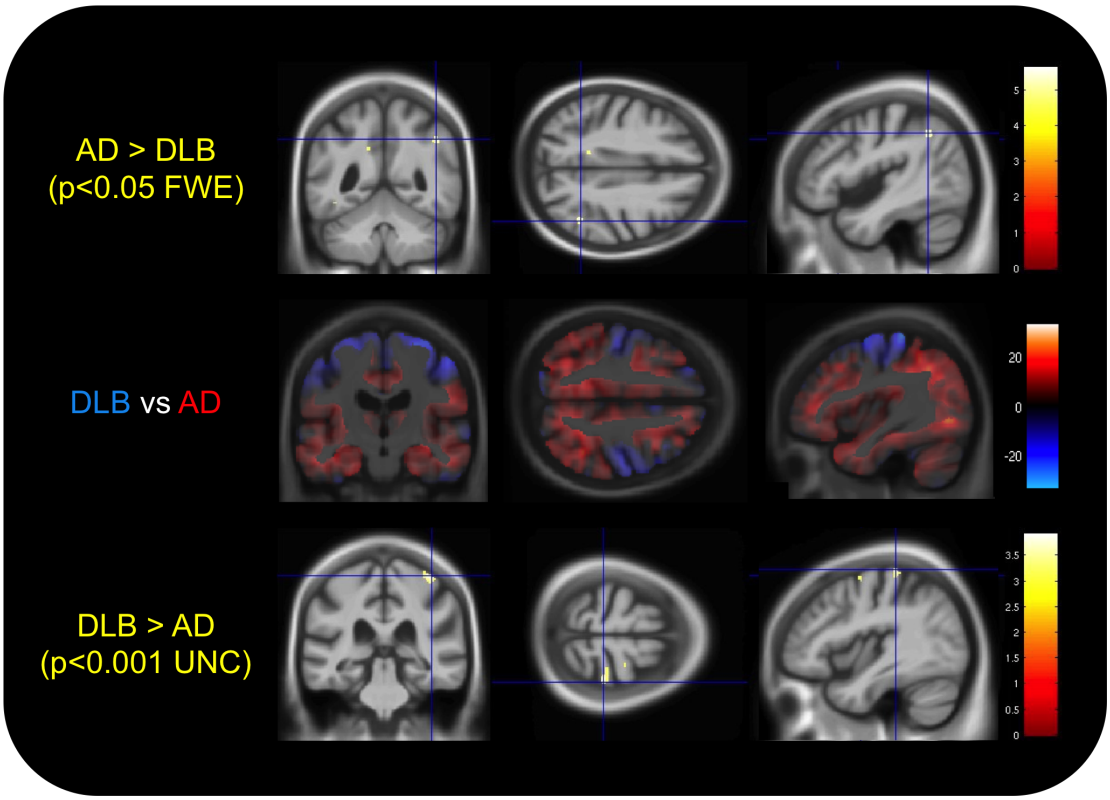


Figure 7.4: The top row demonstrates the regions where volume loss in AD was significantly greater than in DLB. The middle row is the effect size map, with regions in blue more affected in the DLB group, and regions in red more affected in the AD group. The bottom row demonstrates the regions where volume loss was significantly greater in DLB. UNC=uncorrected for multiple comparisons.

effect size map, however, demonstrated a clear anterior-posterior separation between affected areas, with bilateral frontal atrophy in the FTLD-Tau group, and bilateral parietal atrophy in the Alzheimer's disease group (see Figure 7.6).

There was significant grey matter volume loss in left inferior frontal regions and the anterior temporal lobe (p<0.001 FWE) in the FTLD-TDP43 group when compared with the Alzheimer's disease group (see Figure 7.6). The Alzheimer's disease group had greater volume loss in the parietal lobe and around the posterior cingulate sulcus and the parieto-occipital sulcus, but these regions did not survive correction for multiple comparisons. The effect size map shows an anterior-posterior separation of volume loss between the Alzheimer's disease and FTLD-TDP43 group, however, this is less well defined than the previous comparison of Alzheimer's disease with FTLD-Tau, with volume loss in the FTLD-TDP43 group extending further back into posterior temporal

lobe regions. Temporal lobe atrophy was also slightly asymmetric, affecting the left hemisphere more than the right (see Figure 7.6).

7.3.5 Comparison of DLB with FTLD Pathologies

In comparison with the DLB group the FTLD-Tau group demonstrated greater grey matter volume loss in the frontal lobe, with a very similar distribution to the comparison of the Alzheimer's disease group with FTLD-Tau ($p<0.001$ FWE) (see Figure 7.7). Likewise, the FTLD-TDP43 group also demonstrated greater volume loss in the frontal and anterior temporal lobes ($p<0.001$ FWE), although with some extension into the right temporal pole, which was not evident in the previous comparison with the Alzheimer's disease group (see Figure 7.7). The DLB group demonstrated significant volume loss in a very small region in the occipital lobe in comparison with the FTLD-Tau group ($p<0.05$ FWE), and in a small region of the cerebellum in comparison with the FTLD-TDP43 group, however, this did not survive multiple comparisons correction. In both effect size maps, greater volume loss in the DLB group was largely confined to the occipital lobe and cerebellum. The lateral aspects of the pre- and post-central gyri were also involved, although the difference was more pronounced in the comparison with the FTLD-TDP43 group.

7.3.6 Comparison of Tau with TDP43

There were no statistically significant differences in grey matter volume between the FTLD-Tau and the FTLD-TDP43 groups that survived multiple comparisons correction. Uncorrected differences were observed in the frontal lobe, around the anterior cingulate, representing greater volume loss in the FTLD-Tau group; while the FTLD-TDP43 group showed greater volume loss in lateral posterior temporal lobe regions (see Figure 7.5). The effect size map suggests a differential pattern of atrophy with taupoathies associated with more tissue loss in frontal lobe regions, while the FTLD-TDP43 demonstrated greater temporal lobe atrophy, extending towards parietal and occipital lobe regions.

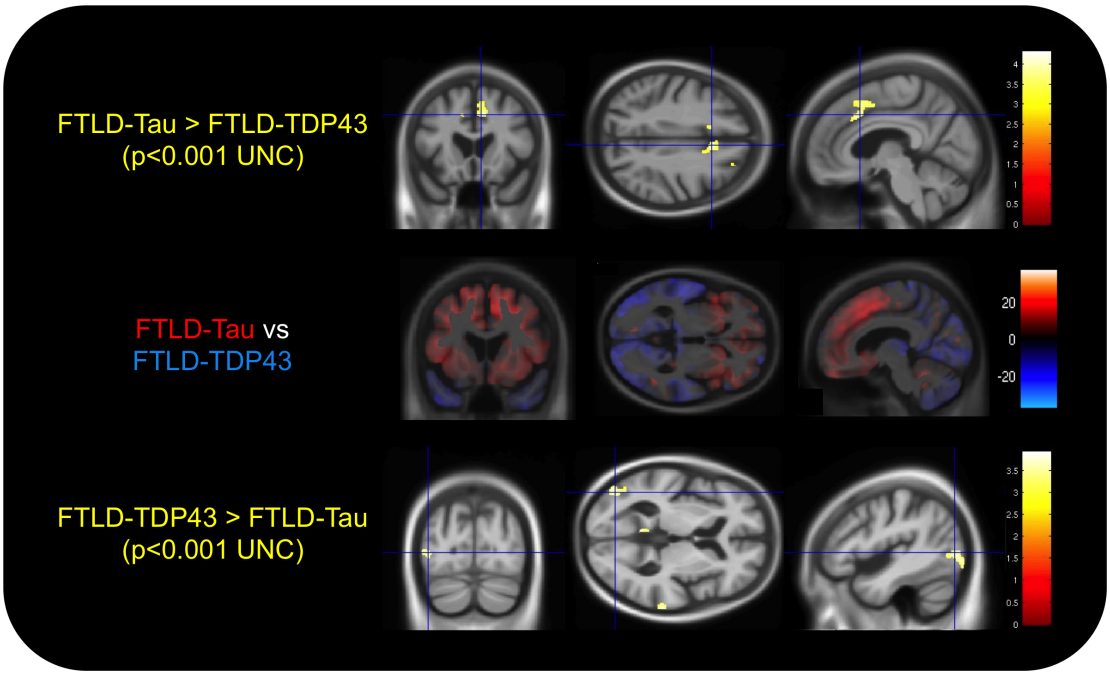


Figure 7.5: The top row demonstrates the regions where volume loss in AD was significantly greater than in DLB. The middle row is the effect size map, with regions in blue more affected in the DLB group, and regions in red more affected in the AD group. The bottom row demonstrates the regions where volume loss was significantly greater in DLB. UNC=uncorrected for multiple comparisons.

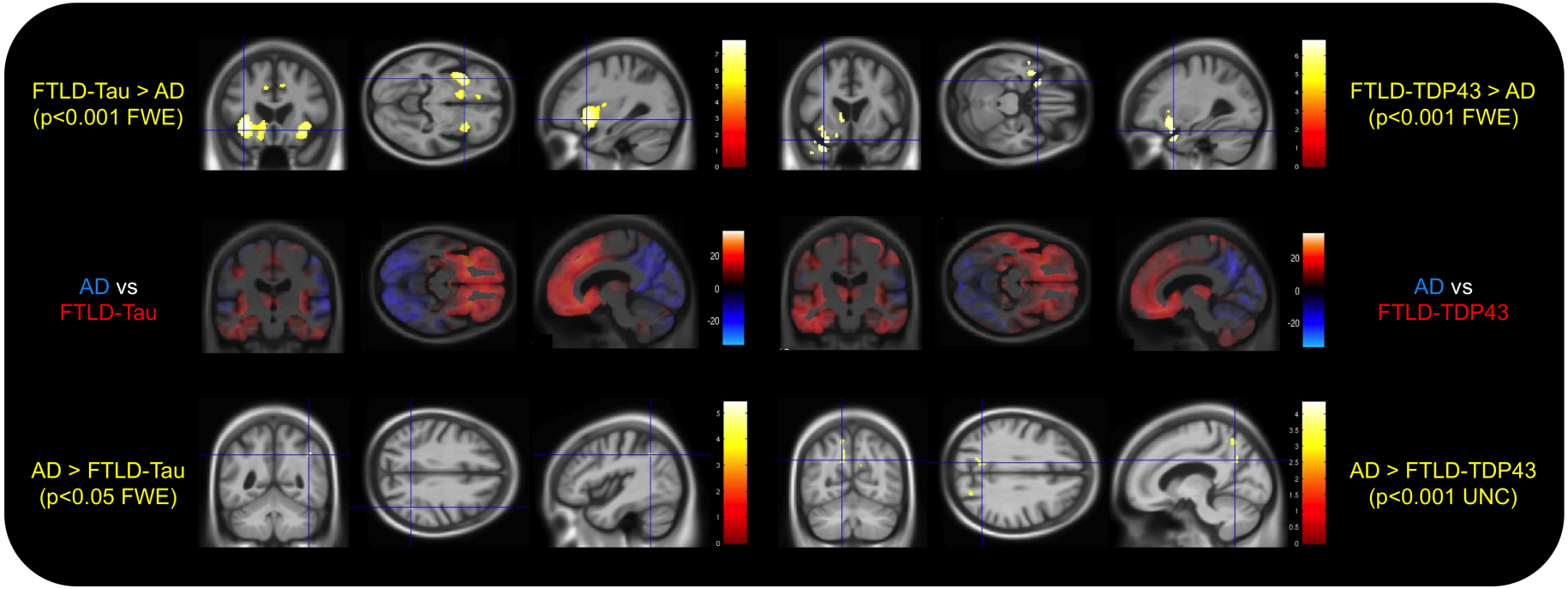


Figure 7.6: Comparison between the FTLD-Tau and the AD group are shown on the left, with the FTLD-TDP43 group and AD on the right. The top row demonstrates the regions where volume loss in FTLD groups was significantly greater than in AD. The middle row is the effect size map, with regions in blue more affected in the AD group, and regions in red more affected in the FTLD group. The bottom row demonstrates the regions where volume loss was significantly greater in AD group than the FTLD groups. UNC=uncorrected for multiple comparisons.

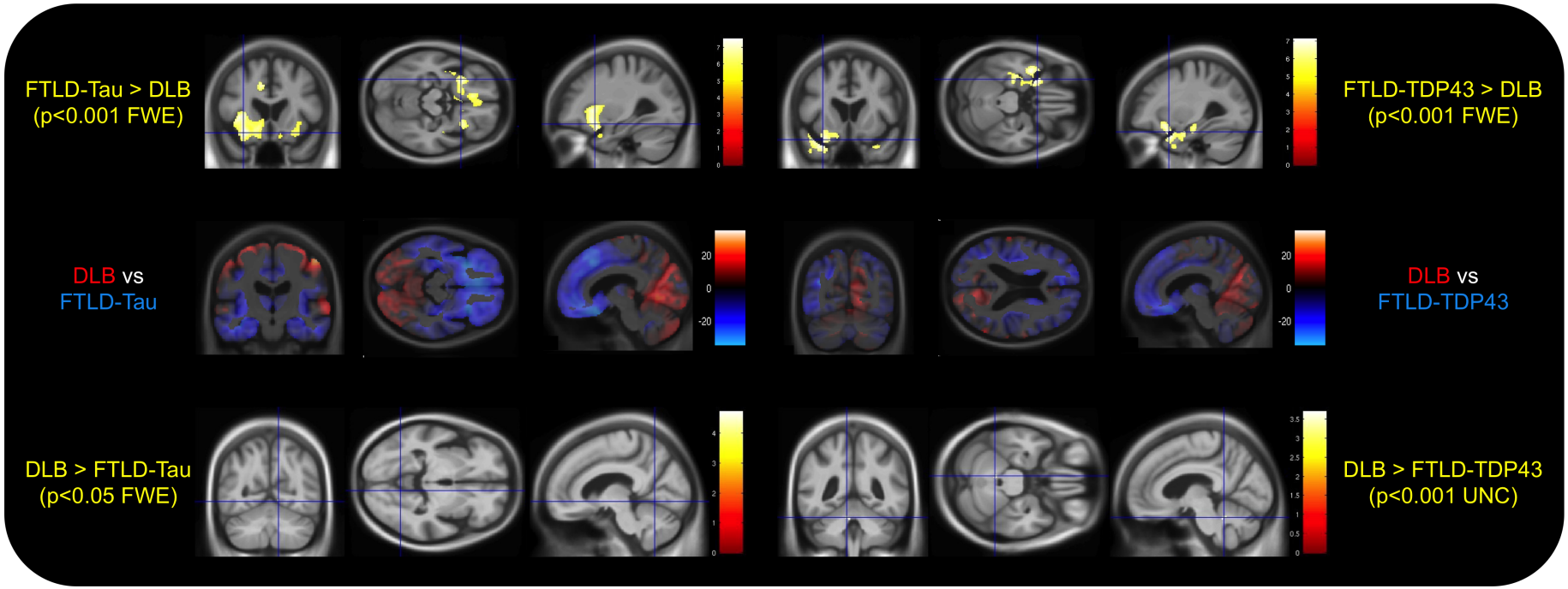


Figure 7.7: Comparison between the FTLD-Tau and the DLB group are shown on the left, with the FTLD-TDP43 group and DLB on the right. The top row demonstrates the regions where volume loss in FTLD groups was significantly greater than in the DLB group. The middle row is the effect size map, with regions in blue more affected in the FTLD group, and regions in red more affected in the DLB group. The bottom row demonstrates the regions where volume loss was significantly greater in DLB group than the FTLD groups. UNC=uncorrected for multiple comparisons.

7.3.7 Comparison of EOAD with LOAD

Although caused by the same underlying pathology, ante mortem differences in grey matter volume were investigated in early and late onset Alzheimer's disease. There were few statistically significant differences between the two groups, however, the EOAD group demonstrated more tissue loss around the parieto-occipital lobe border ($p<0.001$ FWE) (see Figure 7.8). The effect size map demonstrated greater volume loss in the LOAD group in the hippocampus, the occipital lobe, and a very small region around the anterior cingulate gyrus. All other regions were more affected in the EOAD group Figure 7.8).

7.4 Discussion

This VBM study demonstrates patterns of grey matter volume loss associated with the major neurodegenerative dementia pathologies. By combining data from multiple centres it was possible to achieve sample sizes sufficient to estimate consistent patterns of atrophy, not only between patient groups and healthy controls, but also for direct comparison between patient groups.

Using effect size maps to provide an overview of differences in grey matter atrophy between groups highlights the fact that, as well as producing distinct focal patterns of atrophy, neurodegenerative diseases have a more extensive effect on brain tissue. This often goes unreported when focussing only on statistically significant differences, which many studies are underpowered to detect. However, in this study it was also possible to look at statistical differences between disease groups and healthy controls using a very strict statistical significance threshold, controlling the family-wise error rate at a significance level of $p<0.001$. Using these parameters, the Alzheimer's disease pathology group demonstrated an extensive temporoparietal pattern of atrophy, in-keeping with results from previous studies [260, 298, 300]. To look at these changes in more detail, the group was separated into early ($\leq$65 years) and late onset cases, with the LOAD group demonstrating more focal medial temporal lobe atrophy, while the volume loss in the parietal lobe atrophy appeared to driven primarily by the larger EOAD group. Only a small region in the parietal lobe remained significant when comparing

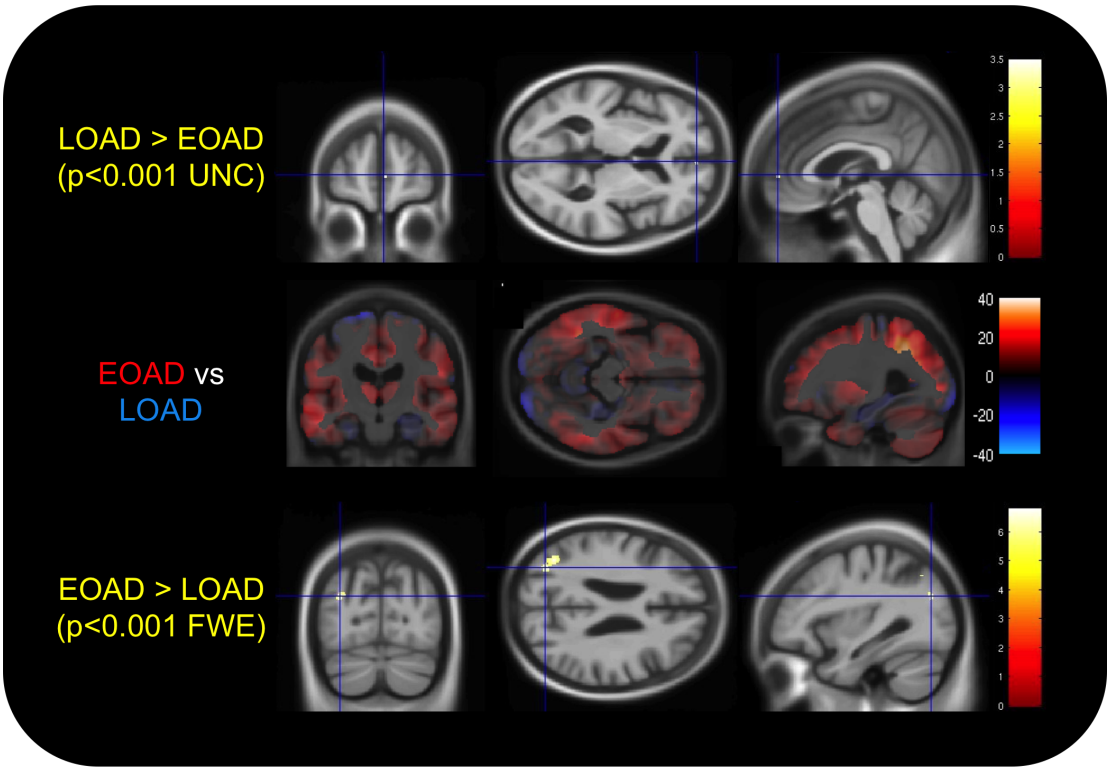


Figure 7.8: The top row demonstrates the regions where volume loss in the LOAD group was significantly greater than in EOAD group. The middle row is the effect size map, with regions in blue more affected in the LOAD group, and regions in red more affected in the EOAD group. The bottom row demonstrates the regions where volume loss was significantly greater in EOAD group. UNC=uncorrected for multiple comparisons.

the groups directly; however, the corresponding effect size map emphasised the medial temporal and parietal separation. As previously suggested in other studies [109, 111, 139], a combination of both medial temporal lobe and parietal lobe atrophy may help to improve the sensitivity and specificity of the diagnosis of Alzheimer's disease.

DLB pathology showed less extensive grey matter loss than the other neurodegenrative pathologies included in this study. Using a very strict statistical threshold, deep grey matter structures such as the thalamus and amygdala, were particularly affected, with relative sparing of the hippocampus, as reported previously by other studies [250, 251, 301]. Direct comparison of the DLB group with both the Alzheimer's disease and FTLD groups using the effect size maps, however, demonstrated relative reduction of grey matter volume around the pre- and post-central gyri in the DLB group, which has also been reported in Parkinson's disease [302] and may reflect more motor involvement in these conditions than in Alzheimer's disease. Comparison with the FTLD

groups also highlighted cerebral atrophy in the occipital lobes, which has previously only been suggested by evidence of hypometabolism using FDG-PET [266].

While there was considerable overlap in the pattern of frontal lobe atrophy between the FTLD-Tau group and the FTLD-TDP43 group, anterior temporal lobe regions appeared to be more involved in FTLD-TDP43 pathologies, particularly in the left hemisphere, while FTLD-Tau pathologies affected more anterior-superior frontal lobe regions. These results should be interpreted with caution however, as they may have been driven by disproportionate representation of TDP43-C pathology in the FTLD-TDP43 group, which usually manifests clinically as semantic dementia, with extensive left anterior temporal lobe atrophy. Larger sample sizes are required to perform more fine-grained analysis between the FTLD-Tau and FTLD-TDP43 subtypes.

While this study benefits from an unprecedented large sample size of ante mortem imaging acquired in patients with pathological confirmation of their diagnosis, and analysis using the most advanced SPM software, it has a number of limitations. Including data from multiple sites increases the sample size and the power to detect statistical differences between groups, however, it also introduces potential confounds relating to the different MRI scanners and pulse sequences used to acquire the images. As the imaging data was acquired over a twenty year period using multiple study protocols, it was not possible to correct for scanner differences to this level of detail, However, an attempt was made to correct for magnetic field strength, which is likely to be the biggest source of variability [303]. Furthermore, although the samples sizes used here are larger than those previously used in the literature, they are not large enough to allow detailed analysis of FTLD-Tau and FTLD-TDP43 pathology subtypes. Similarly, sample sizes were insufficient to stratify groups further by secondary pathology, which may have an important influence on developing patterns of cerebral volume loss. Finally, as with many research studies, there may be selection bias in the patients included in this cohort. Those patients with memory and language led presentations are more likely to take part in research, including brain donation, than patients with more challenging behavioural problems. This may affect the representation of pathologies in this cohort and reduce how well the results generalise to the wider population.

7.5 Conclusions

By combining data from multiple specialist centres, this is the largest study of its kind to attempt to provide comprehensive estimates of volume loss in relation to primary molecular pathology. These patterns of atrophy identified in ante mortem imaging, provide validity for some of features described in Chapter 3, such as the anterior/posterior gradient separating Alzheimer's disease and FTLD; the relative sparing of the hippocampus in DLB vs Alzheimer's disease; and greater asymmetry in the temporal lobe region in FTLD-TDP43 cases compared to FTLD-Tau. In addition, previously unreported findings, such as relatively more atrophy in the pre- and post-central gyri in DLB compared to Alzheimer's disease and FTLD, provide new avenues of investigation for imaging features that could be used to help predict underlying neurodegenerative pathologies during life and improve the accuracy of clinical diagnosis. Further investigation of regional differences using multi-label segmentation propagation will be performed in the next chapter in an attempt to improve the interpretability of these results.

Chapter 8

Differences in Brain Substructure Volumes between Dementia Pathologies

8.1 Introduction

As demonstrated in Chapter 7, the disease processes associated with neurodegenerative dementia lead to volume loss in almost all brain regions, however, the relative severity of tissue loss in certain regions often has predictive value for underlying pathology. While segmenting whole brain images into substructure volumes is anatomically less specific than the voxel-wise approach used in the previous chapter, the results are more easily interpretable and easier to operationalise (e.g. the actual volume of the hippocampus vs the extent of significant voxels in the region) to help distinguish between different neurodegenerative dementias, or to track disease progression. To date, the whole brain, ventricles and hippocampi have been the most common brain structures used for this purpose [111, 249, 251, 304–306]. While other brain regions, such as the thalami [307] and cerebellum [308] have also been investigated, pathological confirmation of underlying pathology is rarely available in these studies and findings are often reported in terms of clinical syndromes.

Manual delineation remains the gold standard segmentation technique, however, it is labour intensive, time consuming and largely impractical for application to multiple brain regions or in large cohort studies. Automated or semi-automated techniques continue to evolve to help address this issue, and advances in basic computing technology may eventually make these a viable option in clinical practice. Multi-atlas segmentation propagation and label fusion may be particularly useful for this purpose due to their high accuracy and ability to segment multiple brain regions simultaneously.

Such techniques (introduced in Section 2.3.3 and described in more detail below) use a library of brain atlases (typically T1-weighted images with corresponding segmentations) together with non-rigid registration to provide a new image with a set of putative segmentations that can then be combined.

In this chapter, the diagnostic utility of applying this approach to a clinically realistic dataset of images is evaluated using a state-of-the-art multi-atlas segmentation propagation and label fusion technique, known as multi-STEPS [309]. The images, acquired during life in a large cohort of pathologically proven dementias, were segmented into 83 brain regions, and the volumes associated with each region incorporated into a SVC and used to predict underlying pathology. In addition, segmentation was also performed using a more detailed multi-label (144 region) atlas library to determine if this increase in detail could improve classification accuracy.

8.2 Methods

8.2.1 Study Population

This study included 181 patients with a diagnosis of dementia during life and post-mortem confirmation of underlying pathology, and 73 healthy control subjects previously described in Chapter 4. Three patients diagnosed with FUSopathies were excluded from the analysis as there was insufficient data to investigate this pathology group. The population consisted of 101 Alzheimer's disease cases, 28 DLB cases, 24 FTLD-tauopathy cases and 28 FTLD-TDP43 proteinopathy cases.

8.2.2 Structural MRI

Image acquisition is described in Section 4.3. All individuals had T1-weighted volumetric MR imaging performed during life.

8.2.3 Template Libraries

Multi-atlas segmentation propagation relies on a library of brain atlases, which are used as templates for segmenting new images. In this study three different template libraries were used, with the first two used to segment brain substructure volumes,

while the third was used to segment whole brain volumes, which were subsequently used to create masks for use in the registration process.

The first template library (known as the Hammers' Atlas) consists of 83 manually segmented regions from 30 healthy volunteers (available from www.brain-development.org. Copyright Imperial College of Science, Technology and Medicine 2007) [310, 311]. The volunteer cohort from which the templates were derived had an equal gender distribution, median age was 31 years, ranging from 20 to 54 years. All scans were acquired on a 1.5T GE scanner using a T1-weighted inversion recovery prepared fast spoiled gradient recall (IR-FSPGR) sequence. A full list of the segmented brain regions is included in Appendix F. Individual segmented regions were also combined to calculate frontal, temporal, parietal, and occipital lobe, the insula and cingulate gyri, central brain structures, ventricles and posterior fossa volumes. A list of the regions contained within each combined region of interest is included in Appendix F.

The second template library consisted of 144 manually segmented regions from 20 healthy volunteers, scanned as part of the Open Access Series of Imaging Studies (OASIS) project (http://www.oasis-brains.org/) [312]. Cortical regions were segmented according to the brainCOLOR protocol (http://braincolor.mindboggle.info/protocols/index.html), while subcortical regions were segmented according to the Neuromorphometrics protocol (http://neuromorphometrics.org:8080/Seg/). A list of the segmented brain regions is included in Appendix G. This cohort included 8 males and 12 females, with an average age of 23.4 years (ranging from 19 to 34 years). Scans were acquired on a 1.5T Siemens scanner using a T1-weighted magnetization prepared rapid gradient-echo (MP-RAGE) sequence.

A third whole brain template library was used to create masks used in the registration and segmentation steps. This library consisted of 839 semi-automated segmentations (described in Freeborough *et al* [275]) of baseline time point T1-weighted images from the ADNI database (www.loni.ucla.edu/ADNI). 682 images were acquired at 1.5T (200 controls, 338 mild cognitive impairment and 144 Alzheimer's disease), while 157 images were acquired at 3T (53 controls, 74 mild cognitive impairment and 30 Alzheimer's disease). This cohort included 474 males and 365 females, with an

average age of 75 years. Additional demographic details are presented in Leung *et al* [313].

8.2.4 Multi-Label Segmentation Propagation and Label Fusion

A brief overview of multi-label segmentation propagation is presented before describing the specifics of the technique used in this study.

Multi-label segmentation propagation and label fusion can be summarised in three stages:

1. The T1-weighted images included in the template library are independently registered to the image under study (the target image).
2. The segmentations associated with the library images are propagated to the target image using the same transformations.
3. The labelled target images are combined (fused) to obtain the final segmentation.

Since some registrations will be more accurate than others, they are typically ranked according to some global or local image similarity measure, and only a subset of the most accurate segmentations are fused into the final segmentation. Label fusion is often based on a voxel level majority voting scheme, however, this method assumes the manual segmentations represent a 'ground truth', error-free segmentation, which is unlikely. To avoid relying on this assumption Leung *et al* adapted the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm [314] for label fusion, developing a probabilistic approach to iteratively construct a consensus segmentation as an estimate of the ground truth, whilst simultaneously comparing this to each transformed atlas as an estimate of image similarity. This technique, known as multiple-atlas propagation and segmentation (MAPS), was found to be highly accurate for segmentation of the hippocampus. Cardoso *et al* later adapted this technique to include local, rather than global, similarity measures (known as Similarity and Truth Estimation for Propagated Segmentations, or STEPS), which improved the accuracy of hippocampal segmentation, and was also adapted for use with multi-label atlases (multi-STEPS). Both the multi-STEPS algorithm and an extension of the MAPS algorithm are used in this study

and will be described below.

As previously discussed in Section 2.3.3, accurate registration of the images under study to the atlases in the template library is essential to allow spatial information to be transferred from one to the other. In this study, using the NiftyReg package (http://sourceforge.net/projects/niftyreg/), each of the atlas images (n=30 for the Hammers' atlas, n=20 for the Neuromorphometrics/BrainCOLOR atlas) was affinely registered (12 DFs) to each target image using a block matching approach [278], and then non-rigidly aligned using a fast free-form registration algorithm [277]. The manual segmentations were then propagated using the previously estimated transformations and resampled using nearest-neighbour interpolation. Label fusion was performed using the multi-STEPS algorithm described above. Within the STAPLE framework, locally normalised cross correlation was applied based on local intensity features, using a Gaussian window, to estimate registration accuracy and rank the atlas images accordingly. A Markov random field model was also incorporated within the framework to add spatial consistency and smooth the borders between the best local classifiers.

8.2.5 Masking

To improve the accuracy of image registration, whole brain masks were constructed for each image in the dataset using the Brain-MAPS technique by Leung *et al*, which is an extension of the (hippocampal) MAPS technique described above for whole brain regions. Image similarity is measured based on the global normalised cross correlation between the target image and the template library of 839 semi-automated segmentations (also described above), and used to rank the registrations from best to worst match after an initial affine registration. Eleven of the best matched atlases (based on previous optimisation analysis) were selected and registered to the target image using affine registration and non-rigid registration based on free form deformation. Label fusion was then performed based on the STAPLE algorithm. The resultant whole brain segmentations were then filled and dilated by 1mm to create a binary mask for use in the registration process.

8.2.6 Quality Control

Quality control based on visual assessment is challenging for such a large number of segmentations and typically only gross misclassifications will be detected. In an effort to make this process slightly more robust, summed regional volumes in eight regions (frontal, temporal, parietal, occipital lobe, ventricles, insula and cingulate gyri, central brain structures and posterior fossa - see Appendix F) and the right and left hippocampi were grouped by primary underlying pathology and plotted against TIV, using the absolute deviation from the median, thresholded at 3.5 times the standard deviation, to highlight outliers. Images highlighted as outliers in any region were then visually inspected in greater detail and excluded if necessary.

8.2.7 Statistical Analysis

Linear SVC were used to predict pathology diagnosis based on segmented brain substructure volumes. Three models were constructed from different feature matrices: the first was based on the 83 segmented volumes in the Hammers' atlas; the second from grouping the 83 features into eight summed regional substructure volumes (frontal, temporal, parietal, occipital lobe, ventricles, insula and cingulate gyri, central brain structures and posterior fossa); and the third using the 144 regional volumes included in the Neuromorphometrics/Brain COLOR atlas. Feature matrices were adjusted for age, gender, TIV and magnetic field strength using a linear model. Fourteen SVCs were used to assess each binary comparison: e.g. the discrimination of Alzheimer's disease pathology from the control group; Alzheimer's disease from DLB; Alzheimer's disease from FTLD-Tau; Alzheimer's disease from FTLD-TDP43; and Alzheimer's disease from DLB+FTLD, etc. Split-half separation was used to divide the data for each classifier into training and testing sets. The training data was scaled to zero mean and unit variance over subjects, with the same transformation then applied to the testing data. SVCs were trained using leave-one-out cross-validation on the testing data and class weighting was applied to adjust for unbalanced groups. The regularisation parameter, C, was optimised using grid-search in the range 1E-5 to 100, increasing by an order of magnitude each time. The SVC was implemented using the squared-hinge loss function and L2 regularisation. Classification accuracy is presented as AUC val-

ues. Feature weighting for each classifier is discussed as an indication of each volume's contribution to group separation [291]. SVC processing and analysis was performed using the Python libraries SciPy 0.14.0 and Scikit-Learn 0.15.2 [286] on Python 2.7.6 64-bit.

SVC performance is reported in terms of sensitivity, specificity, balanced accuracy (0.5*(sensitivity+specificity)) and area under the AUC. Confidence intervals were calculated using the Hanley-McNeil approach evaluated in Newcombe *et al* [290]. Data processing and analysis was performed using Python libraries NumPy 1.8.1, SciPy 0.14.0 and Pandas 0.14.1 on Python 2.7.6 64-bit.

8.3 Results

8.3.1 Demographics

Demographic details for the patients and controls are shown in Table 8.1. Ten controls and four Alzheimer's disease patients were excluded during the quality control process (described below). There were no significant differences between the groups in terms of gender, disease duration and total intracranial volume. The DLB group was significantly older at the onset of clinical symptoms, and at the time of scanning, than the Alzheimer's disease and FTLD-TDP43 groups ($p<0.05$). The control group was also significantly older than the Alzheimer's disease group ($p<0.05$). Time between scan and death was significantly longer for the Alzheimer's disease group compared to the DLB group ($p<0.05$). MMSE within six months of scan date was only available in 112 of the 177 patients included in the analysis (60 Alzheimer's disease, 23 DLB, 17 FTLD-Tau, 12 FTLD-TDP43). Based on the data available, the Alzheimer's disease group had the lowest mean MMSE, significantly lower than the FTLD-tau group ($p<0.05$).

8.3.2 Quality Control

A selection of the quality control plots created for the summed regional volumes and hippocampi are displayed in Figure 8.1, Figure 8.2 and Figure 8.3. Eleven of the 73 controls were highlighted as outliers based on the absolute deviation from the median

	Control	AD	DLB	Tau	TDP43
N	63	97	28	24	28
N at 1.0T	0	7	5	1	7
N at 1.5T	46	83	21	21	18
N at 3.0T	17	7	2	2	3
Gender (%M)	49	60	75	58	50
Age at scan (yrs)	66 (8)	61 (12)	70 (6)	63 (8)	60 (8)
Age at onset (yrs)	-	57 (12)	67 (6)	60 (9)	57 (9)
Disease duration at scan (yrs)	-	3.7 (3.2)	3.0 (2.4)	3.9 (2.1)	3.0 (3.1)
Time until death (yrs)	-	5.6 (3.0)	3.5 (2.3)	5.0 (2.9)	5.6 (3.0)
MMSE (x/30)	-	17 (6)	20 (5)	23 (5)	21 (7)
TIV (mls)	1463 (130)	1470 (144)	1550 (148)	1504 (148)	1482 (155)

Table 8.1: Study population demographics. Results are shown as mean (standard deviation). MMSE is within six months of imaging acquisition and was only available for 112 patients (60 AD, 23 DLB, 17 Tau, 12 TDP43).

in all regions (thresholded at 3.5 times the standard deviation), however, only ten were excluded from the analysis after visual inspection. Only two of these scans had obviously bad segmentations due to white matter disease, while a third had an unusual brain appearance, with a sizeable portion of tissue missing in the fronto-parietal region, with otherwise normal scan appearance. The other seven scans were excluded for having implausibly large hippocampal volumes (>4000mm^3), although these errors were not obvious on visual inspection. Four Alzheimer's disease patients were also excluded due to white matter lesions affecting segmentation accuracy.

8.3.3 Comparison of Regional Volumes

Both raw regional brain volumes and volumes as a percentage of TIV were investigated for group differences. In terms of the raw volumes there were no significant volume differences detected between groups in terms of hippocampi, insula and cingulate gyri, or posterior fossa. In the frontal and temporal lobes, the FTLD groups and the Alzheimer's disease group had significantly smaller volumes than the control group ($p<0.05$). The AD group also had significantly smaller volumes than the DLB group in both the temporal and parietal lobes, and in central structures, and in the occipital lobe when compared to controls ($p<0.05$). Ventricular volume was significantly larger in all disease groups when compared with healthy controls ($p<0.05$). Raw volumes are displayed in Table 8.2.

When considered as a percentage of TIV, left hippocampal volumes were significantly smaller in the patient groups (Alzheimer's disease, DLB and FTLD-TDP43) than in the control subjects ($p<0.05$). There were no significant differences between groups in the right hippocampi or the FTLD-Tau group. In the frontal lobe, temporal lobe, occipital lobe, ventricles and central structures all disease groups had significantly smaller volumes than controls ($p<0.05$). The FTLD-Tau group also had significantly smaller frontal lobe volumes than the AD group ($p<0.05$). In the region of the insula and cingulate gyri, the FTLD groups had significantly smaller volumes than the control group and the Alzheimer's disease group ($p<0.05$). In the parietal lobe the Alzheimer's disease group had significantly smaller volumes than all groups except FTLD-TDP43 ($p<0.05$). In the posterior fossa, the Alzheimer's disease, DLB and FTLD-Tau groups all had sig-

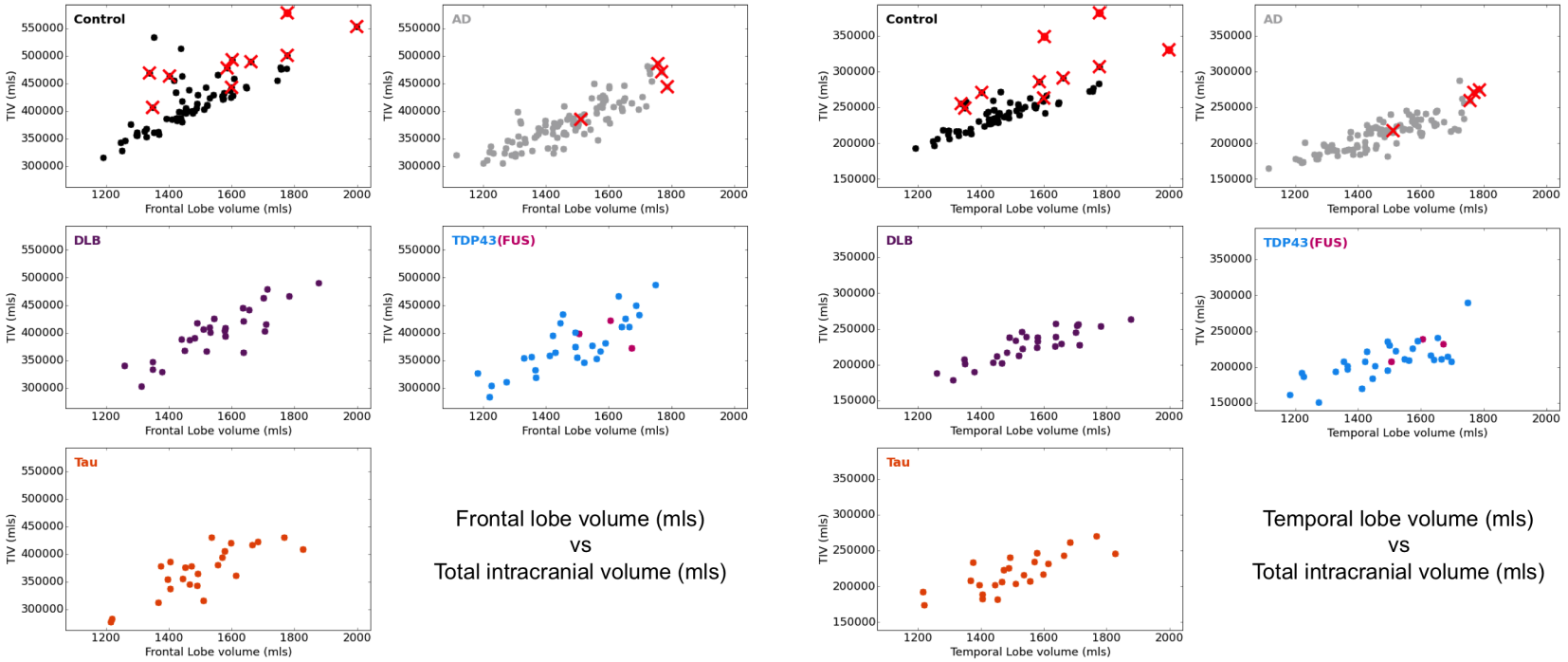


Figure 8.1: Summed frontal and temporal lobe volumes plotted against TIV. Outliers specific to each plot are marked with red dots. A red cross indicates an excluded scan - the outlier status of these scans is not plot specific and may have been flagged by quality control in any of the eight regions. Outliers were detected based on absolute deviation from the median thresholded at 3.5 standard deviations. FUS volumes are shown for interest only and were not included in the analysis.

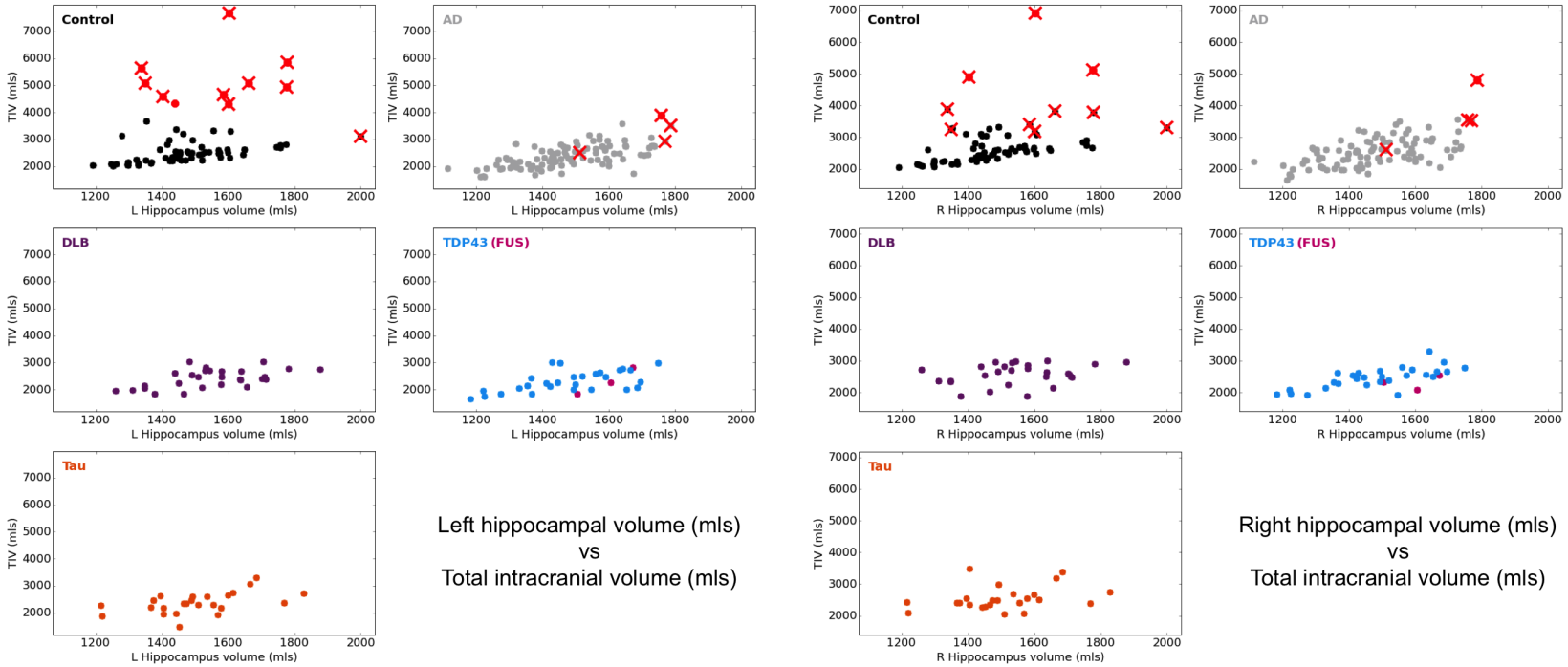


Figure 8.2: Summed right and left hippocampal volumes plotted against TIV. Outliers specific to each plot are marked with red dots. A red cross indicates an excluded scan - the outlier status of these scans is not plot specific and may have been flagged by quality control in any of the eight regions. Outliers were detected based on absolute deviation from the median thresholded at 3.5 standard deviations. FUS volumes are shown for interest only and were not included in the analysis.

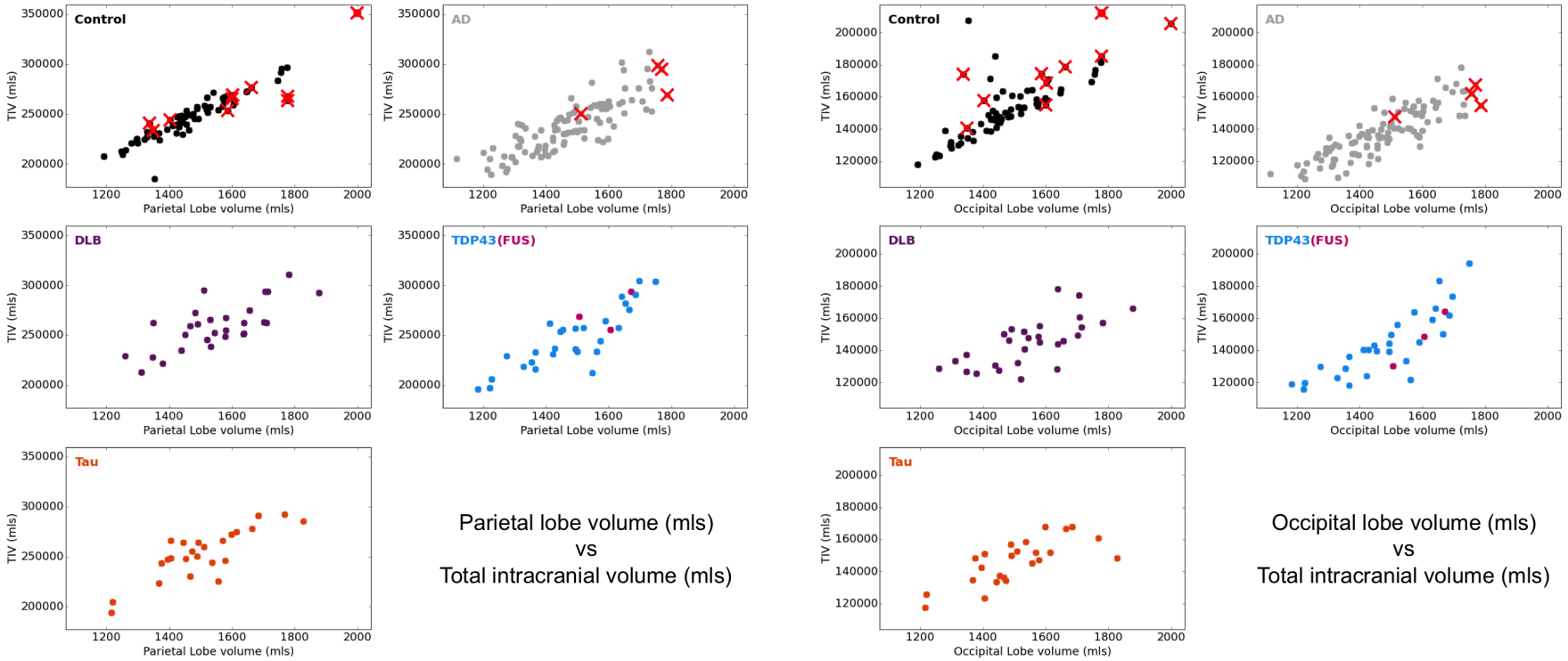


Figure 8.3: Summed parietal and occipital lobe volumes plotted against TIV. Outliers specific to each plot are marked with red dots. A red cross indicates an excluded scan - the outlier status of these scans is not plot specific and may have been flagged by quality control in any of the eight regions. Outliers were detected based on absolute deviation from the median thresholded at 3.5 standard deviations. FUS volumes are shown for interest only and were not included in the analysis.

nificantly smaller volumes than controls ($p<0.05$). Volumes as a percentage of TIV are displayed in Table 8.3.

8.3.4 Diagnostic Accuracy of Support Vector Classifier Using 83 Volumes

The results for each binary SVC, based on the 83 region template atlas, are presented in Table 8.4 and Table 8.5. Classification accuracy of all primary dementia pathologies was high when compared to the healthy controls, although relatively lower for FTLD-TDP43 (AUC=0.84) compared to the other groups (AUC$\geq$0.95). In the head to head disease group comparisons, the SVCs had good accuracy for discriminating Alzheimer's disease from both FTLD pathologies, with slightly higher accuracy in the comparison with FTLD-TDP43 (AUC=0.86) than FTLD-Tau (AUC=0.8). Accuracy for discriminating DLB from FTLD-Tau was also good (AUC=0.79). The most challenging comparisons were between Alzheimer's disease and DLB pathologies, DLB and FTLD-TDP43 pathologies (AUC=0.66 in both cases), and FTLD-Tau and FTLD-TDP43 pathologies (AUC=0.58). Discriminating Alzheimer's disease, DLB or FTLD-TDP43 from all other dementias was also reasonably good (AUC $\geq$0.72). Figure 8.5 and Figure 8.6 present the results of SVC performance in terms of distance plots and receiver-operator curve plots.

8.3.5 Diagnostic Accuracy of Support Vector Classifier Using Grouped Volumes

The results for each binary SVC based on the volumes of the eight summed regions are presented in Table 8.5. Classification accuracy was equal to or lower than the 83 feature analysis in most comparisons, although slight improvements were made in relation to FTLD comparisons, e.g. FTLD-Tau vs FTLD-TDP43 (AUC=0.58->0.71).

8.3.6 Comparison of Segmentation with a 144 Region Brain Atlases

The use of a more detailed atlas library (n=144 regions) produced equivalent classifier performance as the 83 region atlas library in the comparison of disease groups with

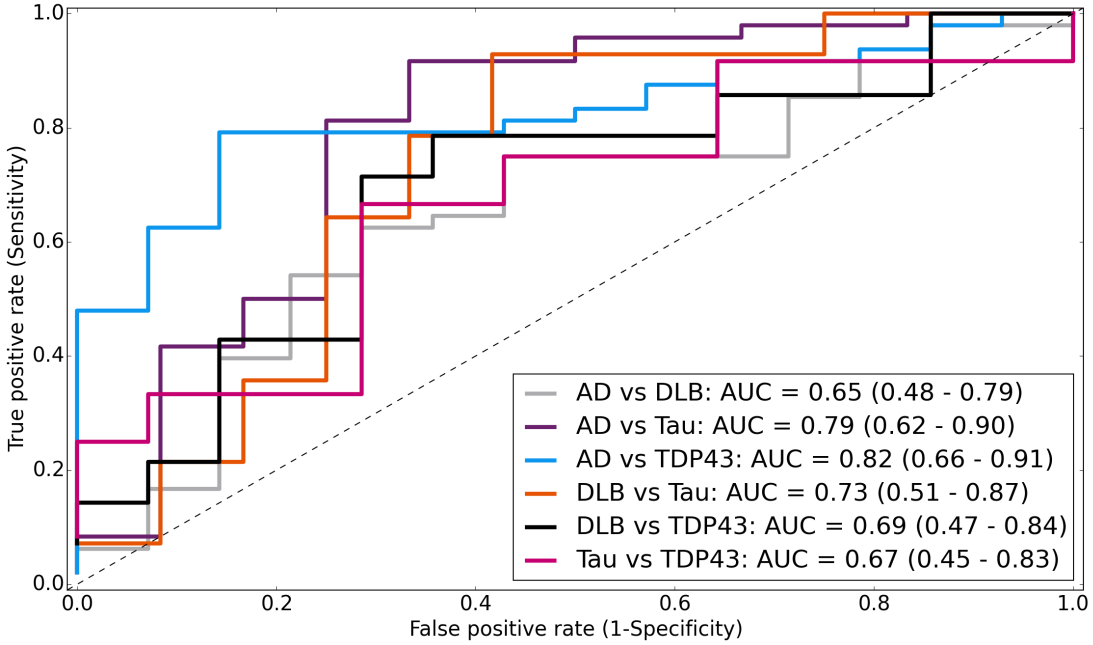


Figure 8.4: ROC plot of SVC performance based on a feature matrix of 144 brain regions. Accuracy is presented as ROC AUC values with 95% confidence intervals shown in brackets

Segmented Volumes (mls)	Control	AD	DLB	Tau	TDP43
Right hippocampus	2.55 (0.32)	2.53 (0.42)	2.58 (0.33)	2.56 (0.38)	2.46 (0.33)
Left hippocampus	2.52 (0.43)	2.36 (0.39)	2.42 (0.34)	2.38 (0.39)	2.31 (0.39)
Frontal Lobe	408 (43.3)	379 (40.8)	400 (46.1)	370 (43.7)	378 (50.4)
Temporal Lobe	239 (20.4)	212 (22.4)	227 (22.7)	218 (25.7)	208 (26.8)
Parietal Lobe	246 (21.6)	237 (25.5)	259 (23.6)	253 (2.49)	247 (30.3)
Occipital Lobe	150 (16.5)	138 (15.4)	145 (14.7)	146 (13.7)	144 (20.4)
Insula & cingulate gyri	69.5 (6.32)	68.4 (7.32)	71.4 (8.50)	67.2 (8.01)	65.8 (8.45)
Central structures	67.3 (7.69)	62.6 (7.61)	66.9 (6.03)	62.9 (5.29)	62.6 (10.1)
Ventricles	24.3 (0.30)	43.7 (12.7)	47.4 (14.4)	44.4 (10.4)	39.1 (11.2)
Posterior Fossa	182 (15.2)	174 (18.4)	180 (19.4)	176 (16.9)	179 (20.9)

Table 8.2: Raw regional volumes for each pathology group presented as mean (standard deviation)

% of TIV	Control	AD	DLB	Tau	TDP43
Right hippocampus	0.17% (0.02%)	0.17% (0.02%)	0.17% (0.02%)	0.17% (0.03%)	0.17% (0.01%)
Left hippocampus	0.17% (0.02%)	0.16% (0.02%)	0.16% (0.02%)	0.16% (0.03%)	0.16% (0.01%)
Frontal Lobe	27.9% (2.2%)	25.8% (1.5%)	25.8% (1.5%)	24.6% (1.8%)	25.5% (1.9%)
Temporal Lobe	16.3% (0.7%)	14.4% (0.8%)	14.6% (0.7%)	14.5% (1.1%)	14.1% (1.3%)
Parietal Lob	16.8% (0.6%)	16.2% (1.0%)	16.8% (1.2%)	16.7% (1.0%)	16.6% (1.0%)
Occipital Lobe	10.3% (0.8%)	9.4% (0.5%)	9.4% (0.7%)	9.7% (0.7%)	9.7% (0.7%)
Insula & cingulate gyri	4.8% (0.1%)	4.7% (0.3%)	4.6% (0.3%)	4.5% (0.3%)	4.4% (0.3%)
Central structures	4.6% (0.4%)	4.3% (0.4%)	4.3% (0.4%)	4.2% (0.3%)	4.2% (0.5%)
Ventricles	1.7% (0.2%)	3.0% (0.8%)	3.0% (0.8%)	3.0% (0.6%)	2.6% (0.7%)
Posterior Fossa	12.5% (0.3%)	11.9% (1.0%)	11.6% (1.0%)	11.7% (0.9%)	12.1% (1.0%)

Table 8.3: Regional brain volumes for each pathology group, as a percentage of TIV, presented as mean (standard deviation)

	Sensitivity	Specificity	Balanced Accuracy
AD vs Control	86% (75%-92%)	94% (85%-97%)	90% (79%-95%)
DLB vs Control	93% (78%-98%)	97% (84%-99%)	95% (81%-99%)
Tau vs Control	92% (75%-97%)	97% (83%-99%)	94% (79%-98%)
TDP43 vs Control	79% (61%-89%)	90% (75%-96%)	84% (67%-93%)
AD vs DLB	67% (49%-80%)	71% (54%-84%)	69% (52%-82%)
AD vs Tau	69% (50%-82%)	83% (66%-92%)	76% (58%-87%)
AD vs TDP43	69% (51%-82%)	86% (70%-93%)	77% (60%-88%)
DLB vs Tau	64% (42%-81%)	100% (84%-100%)	82% (60%-93%)
DLB vs TDP43	79% (57%-90%)	64% (43%-81%)	71% (50%-86%)
Tau vs TDP43	100% (84%-100%)	29% (14%-51%)	64% (42%-81%)
AD vs DLB+FTLD	82% (71%-89%)	59% (47%-69%)	70% (58%-79%)
DLB vs AD+FTLD	57% (41%-72%)	83% (68%-91%)	70% (54%-82%)
Tau vs AD+DLB+TDP43	83% (67%-92%)	49% (32%-65%)	66% (48%-80%)
TDP43 vs AD+DLB+Tau	71% (55%-83%)	79% (63%-89%)	75% (59%-86%)

Table 8.4: SVC accuracy based on the 83 brain substructure volumes in the Hammers' Atlas. 95% confidence intervals are presented in brackets.

	ROC AUC, V=83 (Hammers)	ROC AUC, V=8 (Hammers Grouped)	ROC AUC, V=144 (Neuromorphometrics)
AD vs Control	**0.96 (0.87-0.98)**	0.93 (0.84-0.97)	0.94 (0.84-0.97)
DLB vs Control	**0.96 (0.83-0.99)**	**0.96 (0.83-0.99)**	**0.96 (0.83-0.99)**
Tau vs Control	**0.95 (0.80-0.99)**	**0.95 (0.80-0.99)**	0.94 (0.79-0.99)
TDP43 vs Control	0.84 (0.67-0.93)	**0.87 (0.71-0.95)**	0.85 (0.69-0.93)
AD vs DLB	**0.66 (0.48-0.79)**	0.60 (0.43-0.75)	0.65 (0.48-0.79)
AD vs Tau	**0.80 (0.62-0.90)**	0.70 (0.52-0.83)	0.79 (0.62-0.90)
AD vs TDP43	**0.86 (0.70-0.93)**	0.70 (0.52-0.82)	0.82 (0.66-0.91)
DLB vs Tau	**0.79 (0.57-0.91)**	0.74 (0.52-0.88)	0.73 (0.51-0.87)
DLB vs TDP43	0.66 (0.45-0.82)	0.66 (0.45-0.82)	**0.69 (0.47-0.84)**
Tau vs TDP43	0.58 (0.36-0.76)	**0.71 (0.48-0.86)**	0.67 (0.45-0.83)
AD vs DLB+FTLD	**0.75 (0.64-0.84)**	0.63 (0.51-0.73)	0.59 (0.47-0.70)
DLB vs AD+FTLD	**0.72 (0.56-0.84)**	0.57 (0.41-0.72)	0.64 (0.47-0.77)
Tau vs AD+DLB+TDP43	0.67 (0.50-0.81)	**0.72 (0.54-0.84)**	0.64 (0.47-0.78)
TDP43 vs AD+DLB+Tau	**0.76 (0.60-0.87)**	0.63 (0.47-0.77)	0.76 (0.59-0.86)

Table 8.5: SVC accuracy based on the Hammers' Atlas (83 volumes). Grouped ROC AUC values are calculated based on the 8 (summed) region SVC (frontal, temporal, parietal, occipital lobe, ventricles, insula and cingulate gyri, central brain structures and posterior fossa). Neuromorphometrics/BrainCOLOR atlas accuracy is based on 144 volumes. The highest classifier accuracy for each binary comparison is in bold. 95% confidence intervals are presented in brackets. V = volumes.

	Hammers' Atlas	Neuromorphometrics/BrainCOLOR Atlas
AD vs DLB	Right Parahippocampal and ambient gyri Left Middle and inferior temporal gyrus Left Postcentral gyrus	Right central operculum Left superior frontal gyrus medial segment Left Cerebellum Exterior
AD vs Tau	Left Superior temporal gyrus - anterior part Right Parahippocampal and ambient gyri Left Medial orbital gyrus	Right posterior insula Right Ventral DC Left superior frontal gyrus
AD vs TDP43	Right Substantia nigra Left Inferior frontal gyrus Right Inferiolateral remainder of parietal lobe	Left anterior insula Right opercular part of the inferior frontal gyrus Right cuneus
DLB vs Tau	Right Putamen Left Putamen Left Pallidum	CSF Left inferior temporal gyrus 5th Ventricle
DLB vs TDP43	Left Superior temporal gyrus - posterior part Right Postcentral gyrus Left Middle and inferior temporal gyrus	Right supplementary motor cortex 3rd Ventricle Right postcentral gyrus
Tau vs TDP43	Left Superior temporal gyrus - posterior part Left Middle and inferior temporal gyrus Left Anterior temporal lobe - lateral part	Left posterior insula Right anterior cingulate gyrus Right supplementary motor cortex

Table 8.6: The top three highest weighted features for each classification task, and for each atlas library, are listed as an indication of the regions driving the class separation

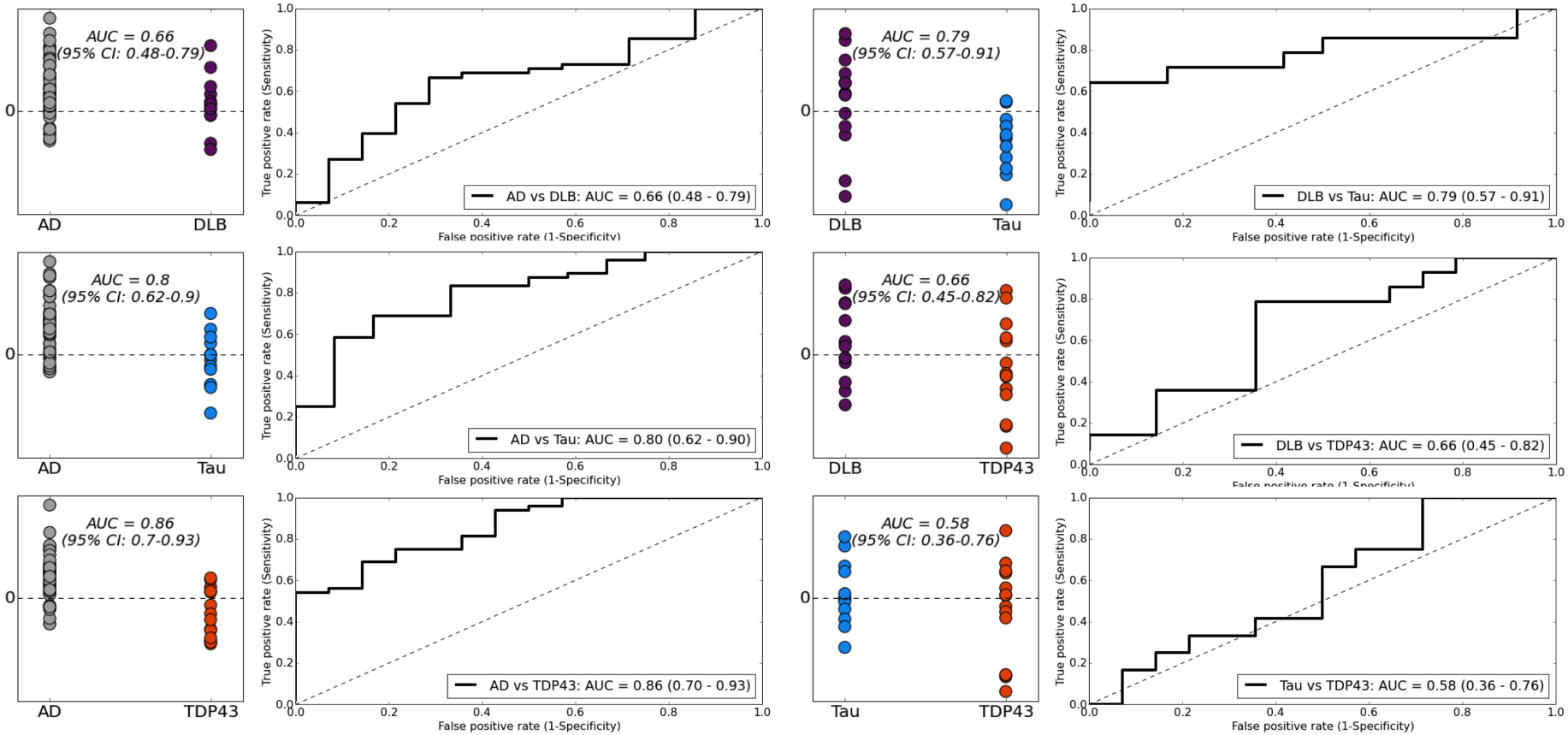


Figure 8.5: Distance plots and ROC plots of SVM performance based on head to head pathology comparisons. Distance plots represent the distance of each image (represented by 83 segmented brain regions) from the separating hyperplane (marked as 0). Performance of each SVM is summarized by the area under the curve (AUC) of the receiver operating characteristic, with 95% confidence intervals (CIs).

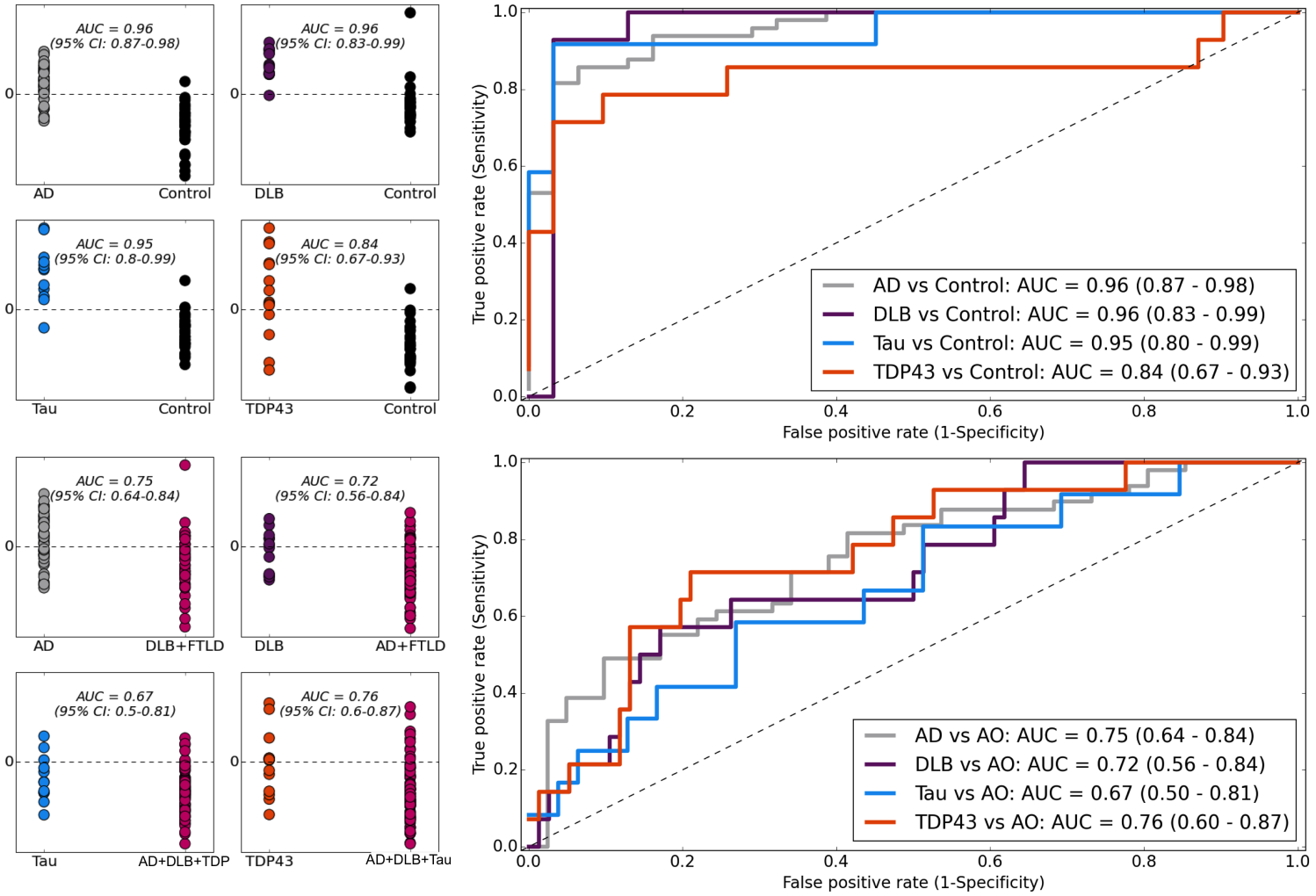


Figure 8.6: Distance plots and ROC plots of SVM performance based on disease groups versus controls, and a single disease group versus all other disease groups combined (AO). Distance plots represent the distance of each image (represented by 83 segmented brain regions) from the separating hyperplane (marked as 0). Performance of each SVM is summarized by the area under the curve (AUC) of the receiver operating characteristic, with 95% confidence intervals (CIs).

healthy controls Table 8.5. Similarly no clear benefits were seen in the head to head disease group comparisons, although improvements were noted in the comparison of DLB vs FTLD-TDP43, and FTLD-Tau vs FTLD-TDP43. Results are presented in Figure 8.4. Table 8.6 provides details of the top three absolute highest weighted features based on each atlas, to provide an estimate of the regions driving the class separation.

8.4 Discussion

In this chapter, the ability to identify underlying neurodegenerative pathology from segmented brain volumes was investigated using a multi-atlas segmentation propagation and fusion technique, and support vector classifiers. Images were segmented into 83 and 144 brain regions respectively using two unique template libraries, these volumes were than arranged into two independent features matrices and used in SVC analysis. Regardless of the feature set used, excellent separation was achieved between patients and healthy controls, while only moderate success was achieved when the disease groups were compared directly. Examination of the feature weights applied in each binary classifier were not always equivalent between the different template libraries (Table 8.6), however, the regions broadly reflected previous reports in the literature, with weighting of the Alzheimer's disease and FTLD-Tau classifier particularly influenced by temporal and frontal regions, while the Alzheimer's disease and FTLD-TDP43 classifier was influenced by frontal and parietal regions. Notably, the post central gyrus, previously shown to be relatively more affected in DLB in the previous chapter, exerted a reasonably high degree of influence over the Alzheimer's disease vs DLB and DLB vs FTLD-TDP43 classifiers.

The 83 region segmentations were also summed according to regional localisation and assessed for group differences using univariate statistics. Analysis of the raw volumes revealed some significant differences, primarily between disease groups and controls, which were in-keeping with previous reports in the literature, and findings from other studies within this thesis. However, performing the same analysis based on volumes as a percentage of TIV produced stronger results, highlighting significant differences between disease groups and healthy controls in almost all regions, as well as some differ-

ences between disease groups. In the frontal lobes for instance, while all disease groups had significantly smaller volumes than controls, significant differences were also detected between the Alzheimer's disease group and the most affected FTLD-Tau group. Both FTLD groups also had significantly smaller volumes than the Alzheimer's disease and control groups in the region of the insula and cingulate gyri, which are central to the salience network thought to be disrupted in FTLD. The Alzheimer's disease group, comprised primarily of early-onset cases (n=69/97), had significantly smaller parietal lobe volumes than all but the FTLD-TDP43 group. The DLB group typically demonstrated the larger volumes in most brain regions compared with the other pathologies, however, there were notable exceptions in the posterior fossa and occipital lobe regions, which have previously been implicated in DLB.

In addition to the larger summed brain regions, hippocampal volumes were also included in this analysis. While there were significant differences between patient groups and controls on the left, there were no differences on the right, and overall the control group demonstrated a relatively small hippocampal volume when compared with previous reports in the literature. This may be attributable in part to a degree of age related atrophy in the control group, who were significantly older than the patients with Alzheimer's disease. Furthermore, since the control cases are not pathologically confirmed pre-symptomatic pathology cannot be excluded. As discussed above, and demonstrated in Figure 8.2, a few implausibly large hippocampal volumes were produced in the control group and were subsequently excluded from the analysis. It is, however, possible that even if they were not as large as predicted by the automated segmentation, they may have been on the large end of the spectrum, and excluding them may have artificially lowered the hippocampal volume range in this group. In terms of comparisons between the disease groups, the raw volumes broadly reflected what might be expected from the literature. For instance, given the disproportionate number of EOAD cases in this cohort, which often have prominent parietal lobe atrophy with relatively spared hippocampi, the FTLD-TDP43 group, with the inclusion of TDP43-C pathology (often severely affecting the left anterior temporal lobe), demonstrated the smallest hippocampal volumes. By comparison, the DLB group had relatively large hippocampal volumes. However, none of these differences could be detected when

considered as a percentage of TIV.

While this study is strengthened by the use of state of the art segmentation techniques and a large, clinically representative imaging dataset of pathologically proven dementias, it has a number of limitations. Primary among these is the difficulty in applying robust quality control measures to the segmented volumes. The inclusion of multiple volumes in the segmentation make accurate visual inspection impractical, if not impossible, and while an attempt has been made here to partially address this issue, it is far from robust. In particular, a large number of healthy control subjects were excluded on the basis of having abnormally large hippocampi, however, the source of this anomaly could not be identified on visual inspection. An alternative practical solution implemented alongside sophisticated segmentation techniques, with the opportunity to adjust or reject poor segmentations, would help to improve the accuracy of the results and provide greater confidence in their interpretation. One possible approach could be the use of one-class SVCs, trained on an existing database of quality controlled scans, which could then be used to assess the quality of new scans prior to segmentation [315]. Quality control issues are potentially amplified in studies of neurodegenerative dementia by the use of template libraries based on young, healthy volunteers, which fail to represent the often complex patterns of atrophy associated with diseased brains. While this issue is beginning to be addressed for hippocampal and whole brain segmentations (see http://cmictig.cs.ucl.ac.uk/niftyweb/), disease specific multi-label atlases are currently unavailable. It should be noted that quality control issues are not unique to multi-atlas segmentation propagation and fusion techniques, and may be underestimated in many brain parcellation techniques. Additional limitations include the imbalance in the pathology groups, disproportionate representation of young-onset Alzheimers disease cases, and failure to take into account secondary pathologies when stratifying the disease groups. As mentioned above, control subjects were not pathologically confirmed, therefore, pre-symptomatic pathology cannot be ruled out in this group, however, this does not affect the more clinically relevant between-pathology group comparisons. Another potential confounding factor particular to this retrospectively compiled cohort, is the variability in image acquisition parameters. While volumes were adjusted for magnetic field strength, there are several other related factors that could bias these results

that it was impossible to correct for. Finally, whilst the sample size is very large in the context of pathologically confirmed dementias, larger numbers in all groups would improve statistical certainty, particularly in the SVC experiments where it is necessary to split the data into training and testing sets.

To date this study is thought to be one of only three studies to apply machine learning algorithms to pathologically proven cases [301, 316], with all three studies achieving similar levels of classification accuracy. The technique applied by Vemuri *et al* [301], based on 91 brain substructure volumes is particularly interesting, and benefits from a training set of "pure" post mortem confirmed cases (without secondary pathologies), whilst also going beyond the binary classification tasks carried out in the other two studies. However, the authors acknowledge they did not have sufficient data to provide an independent testing set, unlike in this study, relying instead on leave-one-out testing. Since pathologically confirmed imaging data is such a rare commodity, only a multi-centre collaborative approach is likely to provide suitable sample sizes to develop these techniques further.

8.5 Conclusions

In conclusion, this study demonstrates the multiple segmented brain substructure volumes, produced using a fully-automated approach, can discriminate between neurodegenerative pathologies. However, the complexity and computational effort used to extract these volumes is not balanced by the level of accuracy achieved in this study, which is comparable with the accuracy achieved using visual rating scales that take, on average, less than three minutes to apply. Improvements in quality control analysis, the availability of disease specific brain atlas templates, and more consistent imaging acquisition have the potential to improve this accuracy. While sharing valuable pathologically proven data between centres would help to achieve sufficient sample sizes for more robust testing.

Chapter 9

Contributions to Open-Science

9.1 Introduction

Open-science is a movement aimed at bringing freedom and transparency to data collection, experimental methodology and analysis in scientific research. The principles of open-science encompass the reusability of scientific data (open-data), the accessibility of scientific communication (open-access), and the sharing of scientific tools (open-research). There is increasing recognition within the scientific community that adopting the principles of open-science is both ethically responsible, helping to make better use of finite resources, and necessary, to accelerate the pace of scientific discovery, and this is reflected by the current requirement for data sharing and data management plans by many funding bodies.

The benefits of open-data, in particular, are relevant to the study neurodegenerative dementia, where long pre-clinical periods, variable disease trajectories, and causation driven by both genetic and environmental factors, complicates data analysis and typically requires large sample sizes to reliably detect group differences, or measurable disease modifying effects. Accordingly, the World Dementia Council (described in Chapter 1) have adopted open-science as one of five priority areas they will champion to stimulate innovation and development of life enhancing drugs, treatments and care for people with, or at risk of dementia [317].

In the absence of definitive biomarkers of neurodegenerative disease pathology, the data at the centre of this thesis is potentially of significant value. Generating a reasonable sample size to accurately stratify primary dementia pathologies by molecular signature, and examine associated ante mortem clinical data for unique signatures of disease, required collaboration and data sharing between three specialist centres. Acknowledging the value of this dataset, the work of this thesis extended to the design and

implementation of a facility to record and share the data between collaborating centres, with a view to sharing it more widely with the research community beyond the lifetime of the project. To raise awareness and engage the general public in dementia research, a citizen science project based on this data was also developed and piloted. Both contributions will be discussed in this chapter. A more detailed discussion of the benefits of open-data is included in Appendix H.

9.2 Clinico-Pathological Database

As previously described, this study is an international collaboration between specialist centres of neurodegeneration in Amsterdam, London and Newcastle. In the first instance, the priority was to build a web-based system to allow remote access to the data from each of the three sites. Initially, an external company (Ixico - http://www.ixico.com/) was contracted to provide this system based on an existing clinical trials management platform, known as Trial Tracker. However, the constraints built into this system, allowing it to adhere to clinical trials governance, made it inflexible to the needs of this project. After eighteen months of negotiation, iteration and testing, the Trial Tracker system was finally abandoned and the decision was taken to build a bespoke application to serve the same purpose. The main steps in this process are outlined below.

9.2.1 Requirements capture

A number of stakeholders across the collaborating centres were consulted to determine the requirements for a sustainable database system that would be useful beyond the scope and lifetime of this project. These included clinicians, psychologists, imaging scientists and pathologists. A governance expert was also consulted to ensure ethical and responsible sharing of data. Following these initial consultations the data itself was interrogated to determine if there was sufficient information available to meet the proposed requirements. Pathological data was interrogated from hard copies of the post mortem report, while clinical details where primarily extracted from the patient's hospital notes. Imaging data was gathered from several sources, and was not always available in the original DICOM format, therefore, tools to convert between the various

formats also had to be sourced as part of the requirements capture process (these tools are described in Chapter 4). As well as data requirements capture, consideration was also given to the practicalities of data entry through discussion with the data manager at the Queen Square Brain Bank.

Based on this initial exercise, the following requirements were formalised:

- Ensure all data is completely annonymised when uploaded to the database. This includes any data flags contained within the imaging headers.
- Ensure the data is secure and the application is protected against common web attacks.
- Provide role specific user accounts with the opportunity to restrict access to some data.
- Avoid the use of free text boxes in data entry forms.
- Provide the opportunity to download data in csv format.
- Provide pre-defined database queries to allow the data to be filtered before download or viewing.
- Allow recording of data from multiple clinical time points, e.g. MMSE, imaging.
- Record first clinical symptom with the opportunity to record all other clinical symptoms.
- Record Alzheimer's disease pathological staging scores with the flexibility to include historical staging schemes.
- Record lifestyle factors with implications for dementia, e.g. smoking, hypertension, diabetes.
- Specifically request data on potentially relevant diagnostic clinical features, e.g. history of seizures or motor features.

9.2.2 Implementation Using a Python Framework

To enable flexibility and the sustained use beyond the lifetime of this project, the data sharing system was implemented using the Django open source web application frame-

work (https://www.djangoproject.com/). The framework is written in Python and encourages component reuse based on modular design. It is, therefore, highly scalable and well suited to a project such as this whose requirements are likely to evolve over time. Additionally, it provides comprehensive security features and a user authentication system to simplify user account management, allowing different levels of access to be set and greater flexibility in how the system can be deployed.

9.2.3 Database Design

To enable database design the data from the London cohort, which constituted over 50% of the available data for this project, was thoroughly interrogated. Relevant data was organised into tables and the relationships between these tables mapped to impose a logical structure on the data (see Figure 9.1). A data dictionary was also created to describe the contents and format of the data stored in each table.

9.2.4 Web-Based Image Viewer

Since the focus of the project is around the imaging data, one of the primary requirements was to allow the scans to be viewed online via a web browser without the need to download the files first. However, few web-based image viewers are available and those that are typically require a commercial license. The Trial Tracker system has a license agreement with an external company to supply web-based imaging viewing capabilities. Since additional license agreements were already in place between Ixico and UCL, the decision was taken to integrate the new Django based platform with Trial Tracker to allow access to the viewer. Clinical and pathological data was stored on the Django system, while imaging data was stored on the Trial Tracker system. Communication between the two systems was achieved using the Hypertext Transfer Protocol (HTTP). From the users perspective, only the Django system was seen. Examples of the web-based forms and viewer are shown in Figure 9.2 and Figure 9.3.

9.2.5 Testing and Deployment

Final deployment of the Django application was to a UCL server already hosting the Trial Tracker system for clinical trial data management. For this reason, live test-

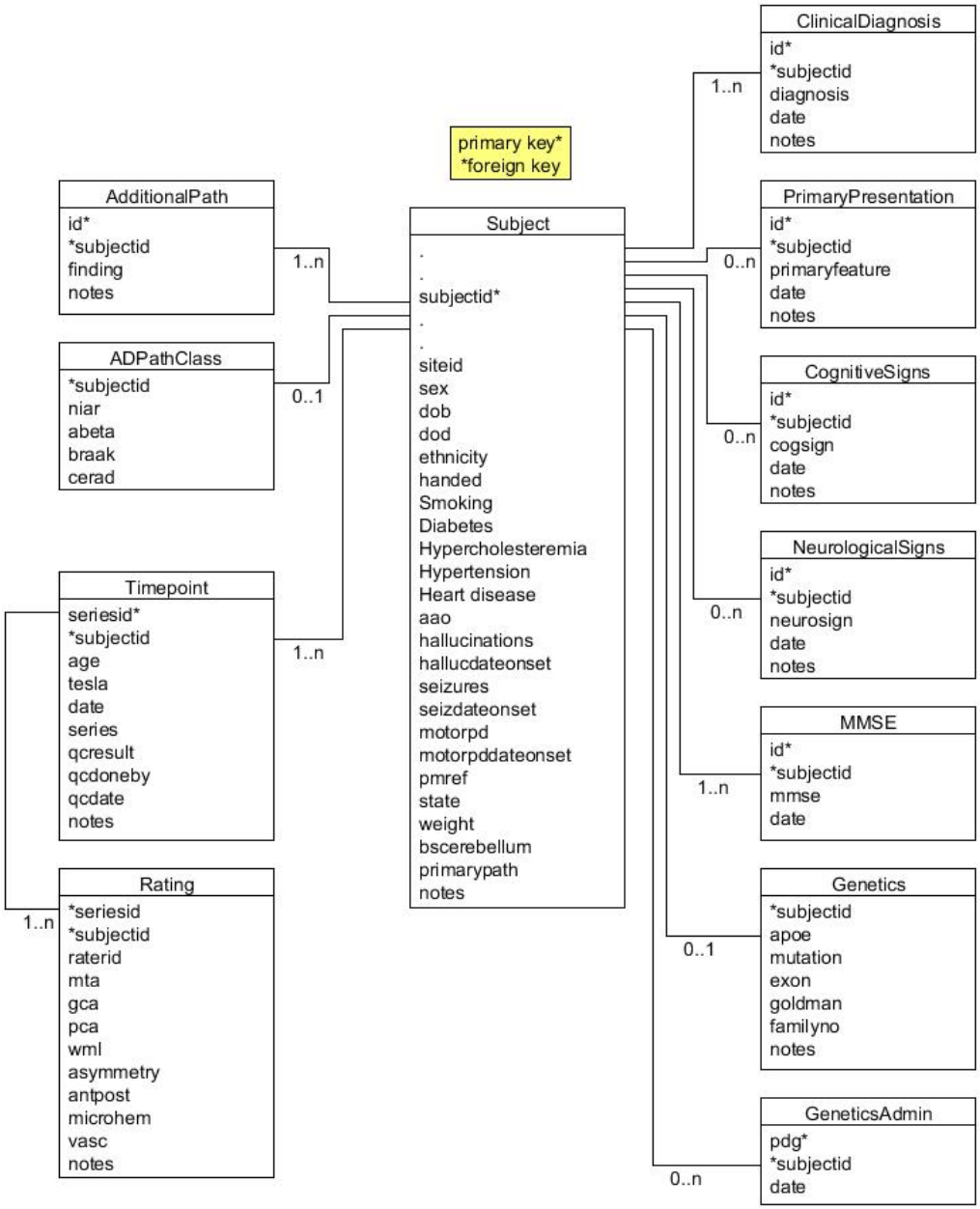


Figure 9.1: UML diagram illustrating the relationship between tables in the clinico-pathological database. Relationships are specified as 0..n: zero to many, 1..n: one to many, 0..1: zero to one. An additional table (Rating) is included for future applications (described below)

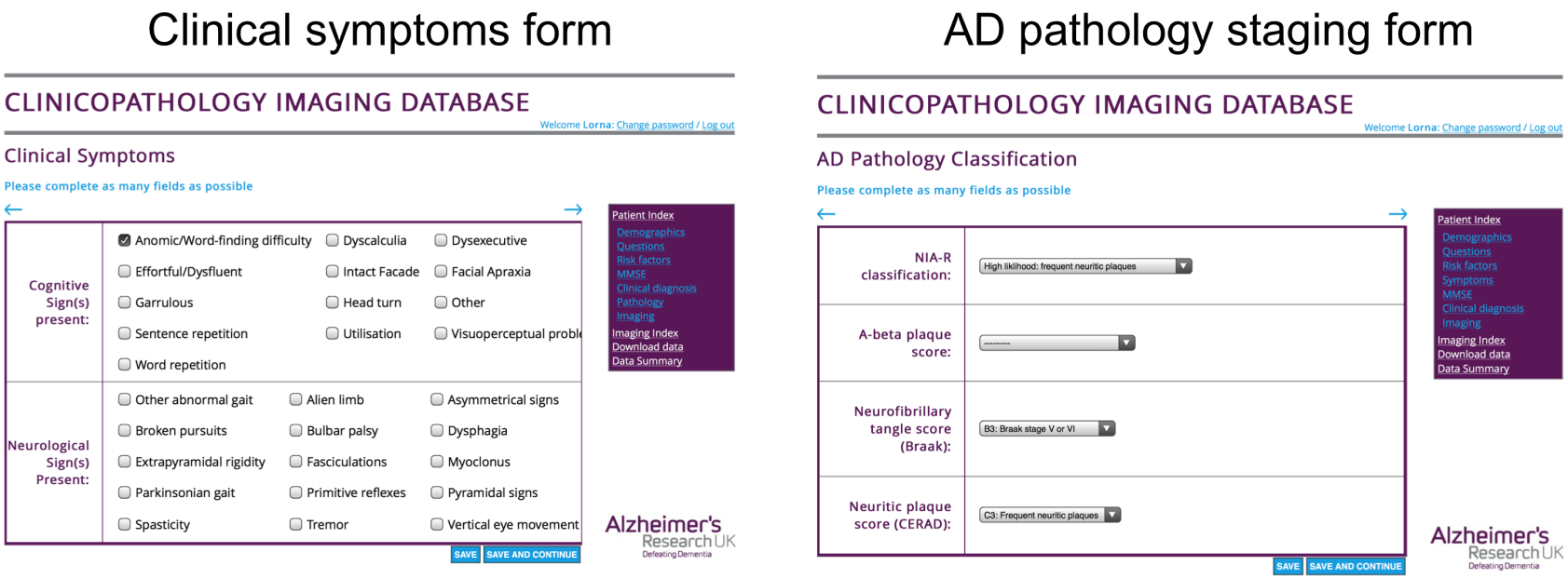


Figure 9.2: Image viewer associated with the bespoke clinicopathological database application and the patient index page.

Subject index

CLINICOPATHOLOGY IMAGING DATABASE

Welcome **Lorna**: Change password / Log out

Subject Index

Click on a Subject ID to edit details

Subject ID	Site	Age at onset	Clinical Diagnosis	Primary Pathology
2402110088	London		Alzheimer's disease	DLB: Lewy body pathology
2403140191	London	NA	No degenerative brain disease	Normal brain
2403310124	London	70	Alzheimer's disease	AD neuropathological changes
2404210147	London	71	Alzheimer's disease	AD neuropathological changes
2408190166	London	69	Alzheimer's disease	DLB: Lewy body pathology
2412280005	London	60	Alzheimer's disease	AD neuropathological changes
2504230146	London	71	Alzheimer's disease	AD neuropathological changes
2507270151	London		Alzheimer's disease	AD neuropathological changes
2509120067	London	65	Posterior cortical atrophy	AD neuropathological changes
2510180071	London	69	Behavioural variant FTD	Tau CBD
2602030059	London		Alzheimer's disease	AD neuropathological changes
2602100126	London	67	Semantic dementia	TDP43C
2605060039	London	71	Vascular cognitive impairment	Small vessel disease
2605290006	London	65	Behavioural variant FTD	TDP43A
2606240154	London	62	Progressive supranuclear palsy	Tau PSP
2607190134	London	68	Posterior cortical atrophy	AD neuropathological changes
2608210140	London	67	Semantic dementia	TDP43C
2612280107	London	71	Semantic dementia	TDP43C
2703030104	London	63	Semantic dementia	TDP43C

Add new subject
Imaging Index
Download data
Data Summary

Alzheimer's Research UK
Defeating Dementia

Web-based image viewer

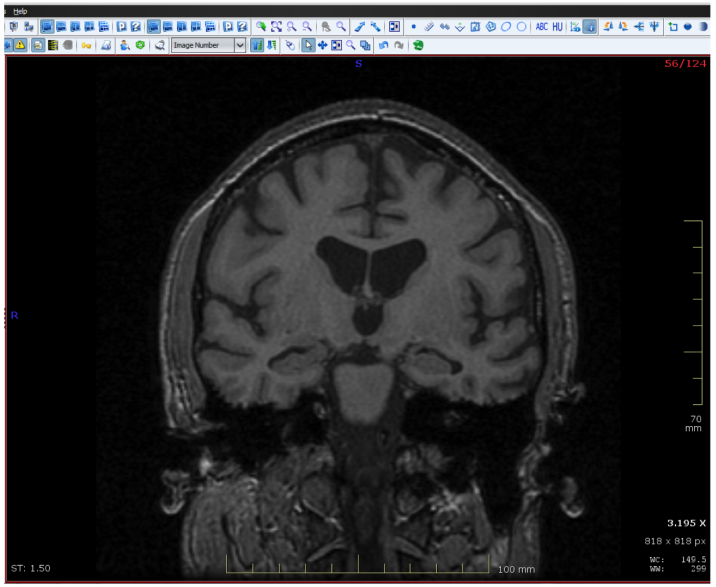


Figure 9.3: Example data entry forms from the bespoke clinicopathological database application.

ing on this server was not possible. To get round this issue, Amazon Web Services (http://aws.amazon.com/) was used to set up a server mirroring the operating system (CentOS) and web server software (Apache) installed on the UCL server. With this set up the database system was deployed and tested for data entry, site navigation and data download. After satisfactory testing the system was successfully transferred to the UCL server.

9.2.6 Future Work

The bespoke web-based database application described here was developed to store the clinical, imaging and pathology data available from three specialist centres in neurodegeneration. It was built to be flexible and sustainable beyond the lifetime of this project and provides a good foundation to expand this collaboration to other centres, allowing this valuable dataset to be shared among the wider scientific community. Several good open-source NIfTI viewers are also now available, allowing the database application to be uncoupled from the Trial Tracker system and eliminating the need for an expensive license agreement. Sophisticated user account management provides the opportunity to limit access to the data, allowing the system to be developed for other purposes such as teaching and training. In the first instance, this application would work well to teach clinicians or researchers to recognise characteristic patterns of atrophy associated with neurodegenerative dementia. Furthermore, it could also be used to train people to apply visual rating scales, and could subsequently be used for refresher training or audit purposes.

9.3 Citizen Science

Citizen science is defined as scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions. It provides opportunities for scientists to engage with the public, and allows the public to play a unique role in scientific research. Inspired by the successes of Cancer Research UK to harness this approach, the Dementia Innovation Unit, formed as part of the World Dementia Council, have begun to explore similar ideas in relation to dementia research.

Parallel to this, the launch of the Join Dementia Research (JDR) Initiative (https://www.joindementiaresearch.nihr.ac.uk/), which provides a public facing portal for recruitment into dementia research, has produced a large community of volunteers (many of them healthy controls) keen to take part in research. This unique community are potentially well suited to trial proof-of-principle citizen science studies; while these studies provide a means of keeping the community engaged until they are recruited in to more conventional research.

Having been seconded to the Dementia Innovation Unit during my PhD, I was asked to be part of the World Dementia Council's citizen science steering committee. As an extension to this work I was asked by the Office of National Director for Dementia Research, to develop a proof of principle citizen science project for JDR. This became a collaborative project with Dr Ferran Prados Carrasco (Translational Imaging Group, UCL). The development of this project, and the learning points that have arisen so far, are discussed below.

9.3.1 Project Specifications

As a largely untested approach to public engagement/research in the context of dementia, and given the infancy of the JDR initiative, tight constraints were imposed on this project. The primary constraints are outlined below.

- To develop an online application to engage the healthy control volunteers already enrolled in JDR. However, affected JDR volunteers should not be excluded from the project.
- The application should be stand alone, unconnected to the JDR platform, and unable to extract any of the volunteer data it contains.
- Minimal personal/demographic data should be collected to prevent overburdening volunteers who have already submitted details to JDR.
- The project should be simple but scalable, with an initial focus on raising awareness/engagement, and gauging the response to this type of approach.
- Citizen scientists should actively generate data through participation, rather than passively donate data as is typical of more conventional research studies.

- The project should be delivered within two months to coincide with the national launch of the JDR initiative.
- Given these time constraints, ethics approval must already be in place for the planned study.

9.3.2 Concept Development

While project ideas were requested from a number of sources, a simple visual image assessment application was thought to be the most straightforward to produce based on the constraints outlined above. After seeking advice from Alzheimer's Research UK on public engagement, this motivation was strengthened by their observation that the public often fail to perceive dementia as a brain disease, and rather as a condition that is "all in the mind". By allowing the public to actually see images of the brain and the effects of dementia pathology on brain tissue, alongside images from healthy controls, it was thought this project could help to alter that perception.

Based on the requirement for existing ethics approval, the imaging dataset collated as part of this thesis, with blanket coverage to use and share it for any research purpose, was thought to be the most appropriate. However, consultation with local ethics experts was also sought to clarify this position.

Looking to other citizen science projects for inspiration (particularly cancer research), it was initially thought that the citizen science task at the centre of the application should be "gamified" to make it more appealing. However, after discussion with Prof. Chris Lintott, co-founder of Galaxy Zoo (one of the most successful citizen science projects to date), and primary investigator for Zooniverse (the world's largest citizen science portal with over one million users), this idea was dropped. Prof. Lintott advised that in his experience, citizen scientists preferred an authentic experience, reflecting a professional scientist's view of the data, than an artificial "gamified" environment.

Based on this advise, and building on the work presented in Chapter 6, the concept for this citizen science project was based on the well established Scheltens' medial temporal lobe atrophy scale, described in Chapter 2. Citizen scientists would be presented with images of the medial temporal lobe, and based on reference images for each scale

point, asked to apply the MTA scale separately to the left and right hemispheres.

From a scientific perspective, and in line with the conclusions from Chapter 6, the motivation behind this study was to investigate if completely naive raters could reliably apply the MTA scale. Proving this to be the case, along with the findings presented in Chapter 6, could help to encourage greater uptake among medical professionals, and in turn add diagnostic value to the sometimes vague radiology reports generated in the investigation of cognitive impairment. However, equally motivating in this initial pilot study was to gauge how the JDR community would respond to this type of approach. From the point of view of the citizen scientist, it was hoped that this study would provide greater understanding of the effects of neurodegenerative disease on the brain, and a rewarding experience of participating in dementia research.

9.3.3 Seeing Dementia

The "Seeing Dementia" project, as it became known, is based on the 80 image subset (20 Alzheimer's disease, 20 DLB, 20 FTLD, 20 controls) rated by the visual rating experts, outlined in Chapter 6, with the expert ratings used here as a gold standard. As a measure of reliability, citizen scientists will be compared to the expert ratings, and to each other. Mean citizen scientist ratings will also be used to train and test a SVC and iteratively determine the number of raters required to achieve maximum classification accuracy. The data used to generate the optimal classifier performance could then be used as a training set against which any new data could be tested.

The application was written in PHP based on a PostgreSQL database, with a few additional features implemented in JavaScript. Dr Ferran Prados Carrasco was the primary developer in this respect. I developed site content and "look and feel" using HTML and CSS within the Bootstrap framework, to provide dynamic layout adjustment between devices (e.g. desktop, tablet, mobile).

In terms of content, both an introductory video and a tutorial video were created using the Camtasia software (see Figure 9.4). Informative content relating to dementia, MRI brain imaging, the hippocampi and the medial temporal lobe scale were all generated to provide users with additional learning opportunities (see Figure 9.5). In terms of the main application, a single, optimal, medial temporal lobe slice is displayed to the

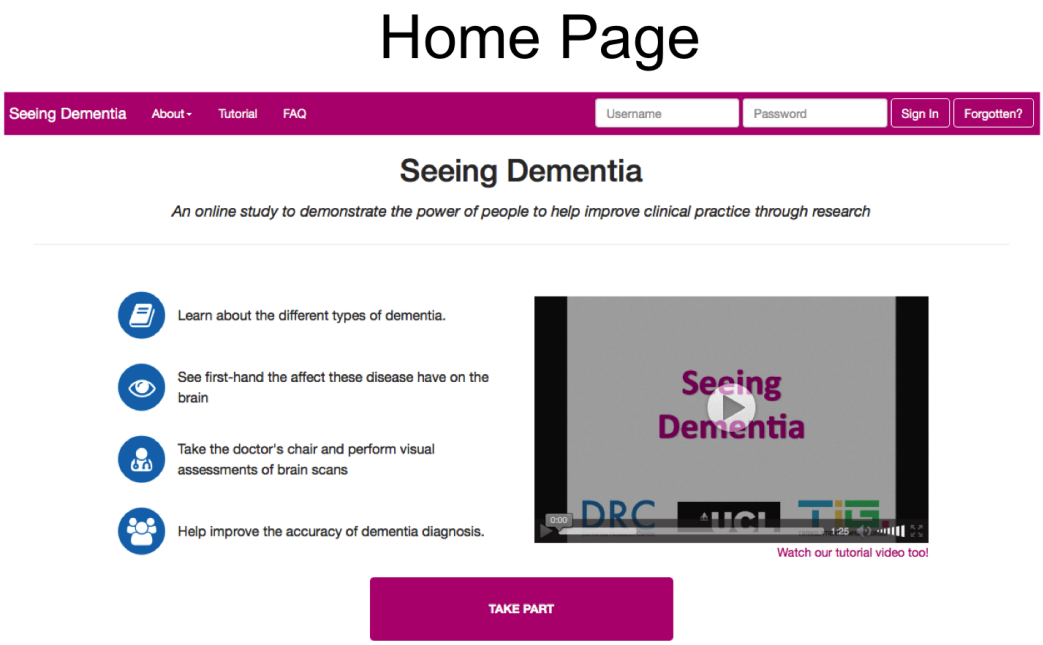
user. Sections of the image can be magnified by hovering over it. Five image buttons are displayed below, representing the medial temporal lobe scores from 0—4 and the user is asked to select the button that most closely matches the MTL shown in the main image (see Figure 9.6). If the right medial temporal lobe (MTL) is shown on the button, the user should score the right hemisphere, and vice versa. To break up the task users are promoted in 5 stages from PhD student (after 20 ratings) to professor (if all 160 ratings is complete) (see Figure 9.6). With no requirement for citizen scientists to rate every scan, images are presented at random to provide a more even distribution of ratings across the 80-image scan subset. In terms of user information, only gender and age range are requested, however, the user is asked to provide an email address and password to allow them to log in and out of the application, and start from the scan where they left off in the previous session (see Figure 9.4). To allow collection of this information the project was registered with the UCL Data Controller. Application was also granted by UCL for a project specific web (http://seeingdementia.ucl.ac.uk) and email address (seeingdementia@ucl.ac.uk). Google Analytics was embedded within each page to provide additional details about how users interact with the project.

9.3.4 Initial Results

To date, a pilot study of the Seeing Dementia project has been carried out as part of the iterative design process. A subset of the JDR community registered in the Yorkshire area were targeted for feedback. Yorkshire was chosen at random from the regional areas in the UK with the fewest JDR registered studies at that time. The primary findings from this exercise are described below. Accuracy of citizen scientist ratings is not reported during this development phase.

User Demographics

547 people were registered with JDR within the Yorkshire area. Based on previous experience, around 10% of registered users who match to a study will participate. Within four weeks 94 people (17%) had registered with Seeing Dementia, however, only 51 (9%) of those became active users within that time. Figure 9.7 demonstrates the breakdown of users by age range and gender. 80% of registered users and 82% of active users were female. The majority of users were middle aged. Users within the 70-79 year old



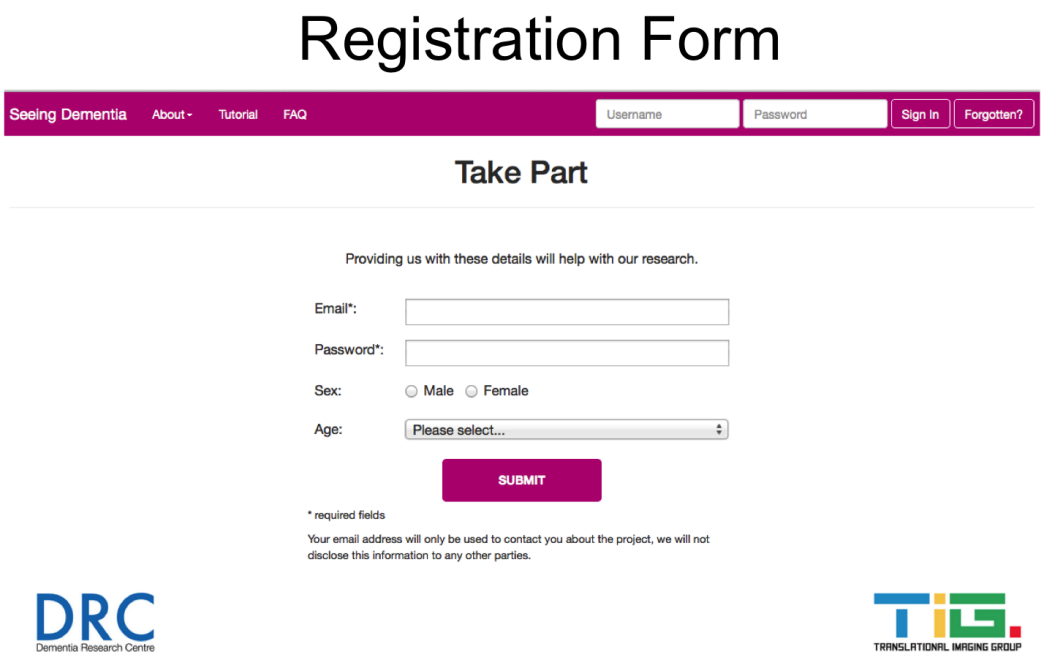


Figure 9.4: Screen shots of the home page and registration form.

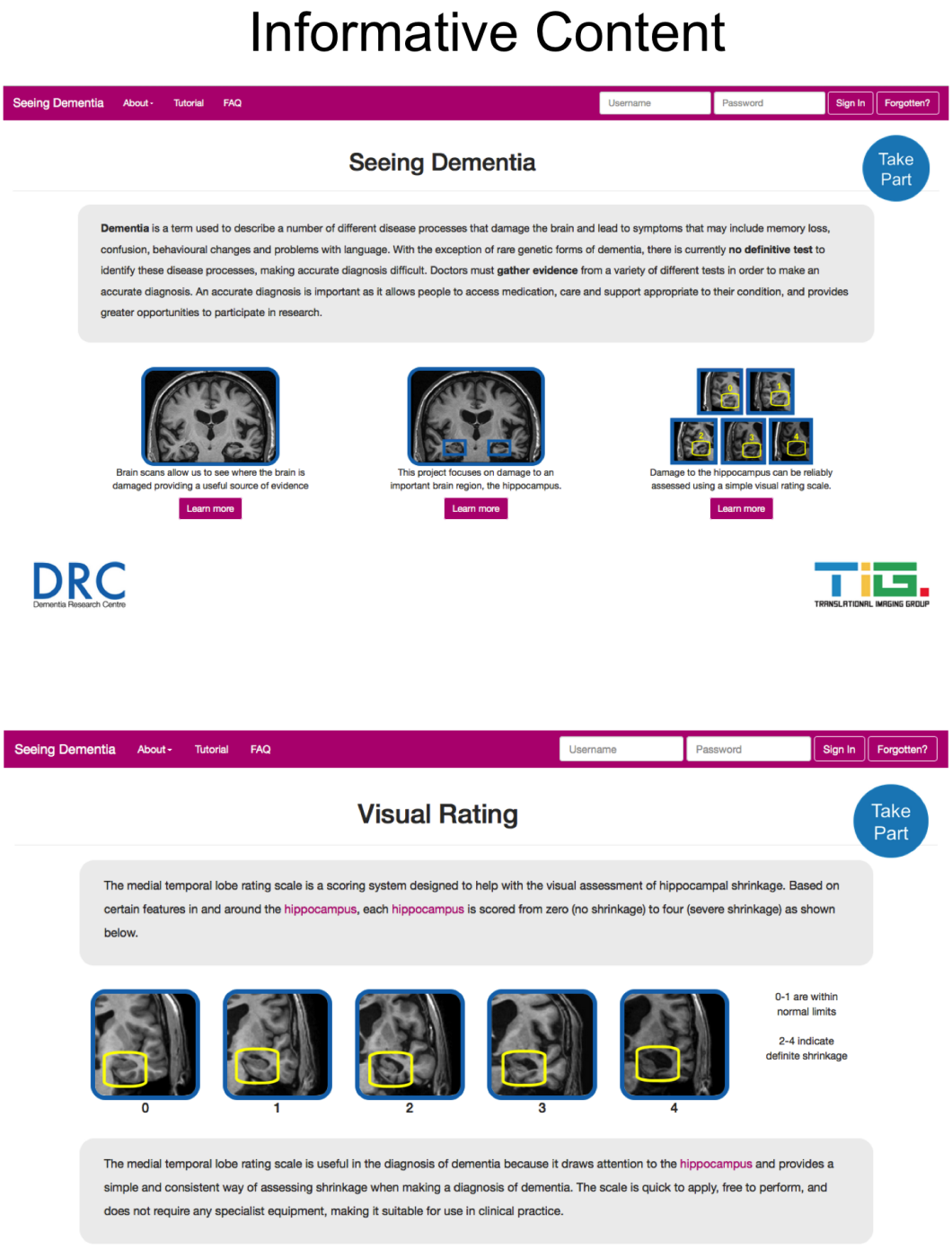


Figure 9.5: Examples of the informative content included with Seeing Dementia.

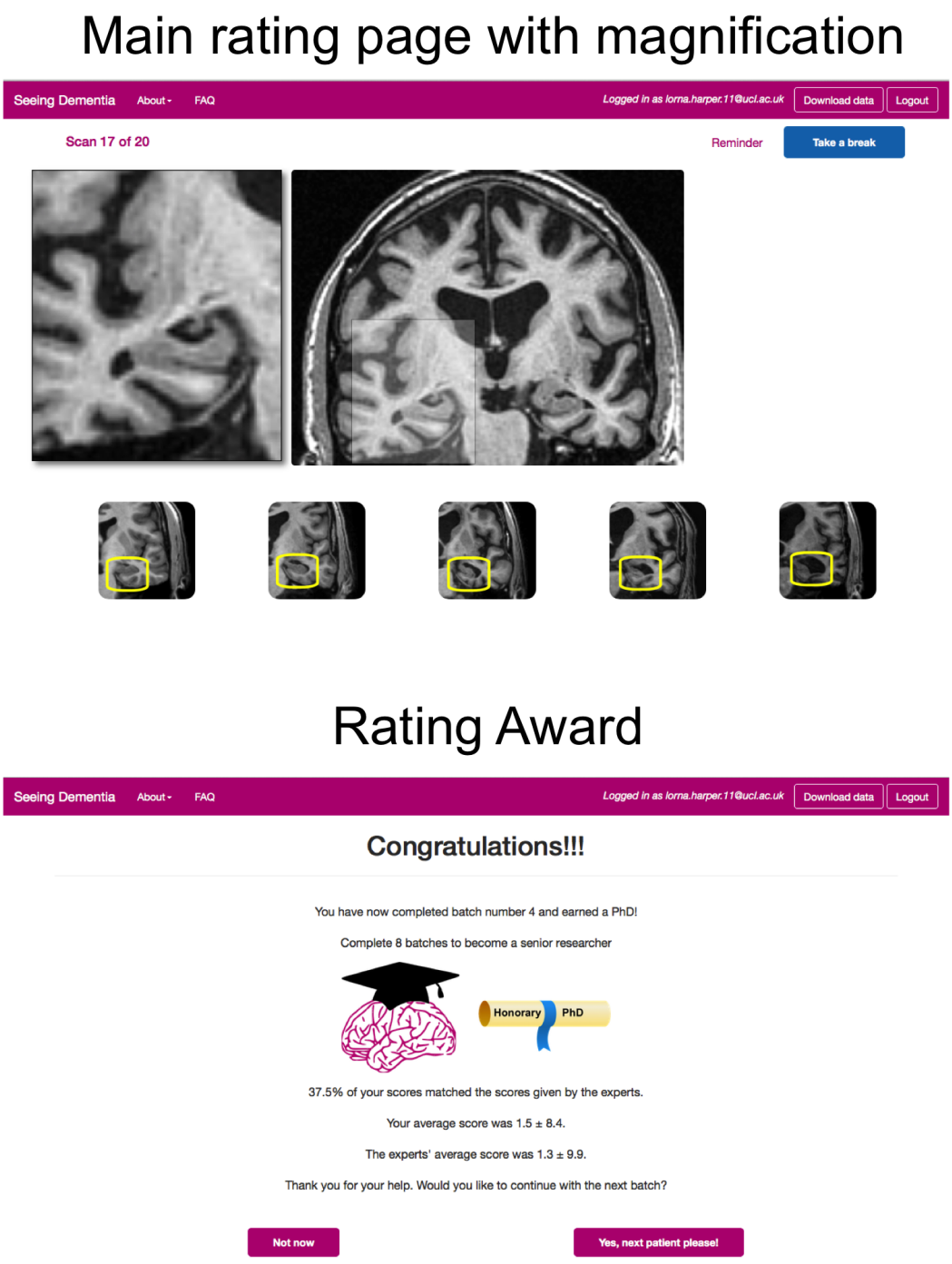


Figure 9.6: Main rating page shown with magnification. Buttons reflecting the MTA scale points are shown below the main image. After a predefined number of ratings, the citizen scientist is promoted. The user is given some feedback on how they compared to the experts based on the percentage of matched scores.

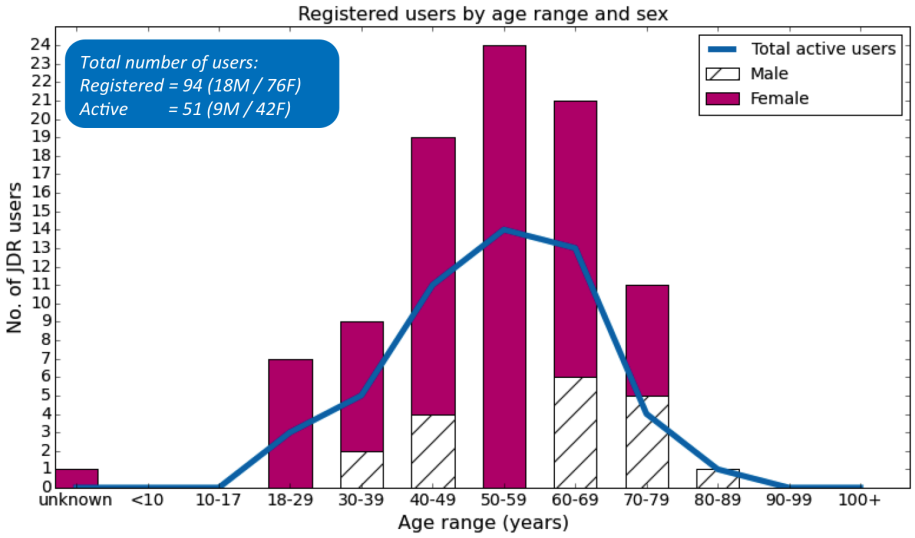


Figure 9.7: Seeing Dementia user demographics. The bars reflect the registered users while the blue line reflects the active users.

bracket were had the lowest proportion of active users to registered users. One female user declined to disclose her age. Due to limitations of the JDR system it is currently not possible to automatically determine how accurately this reflects the Yorkshire cohort overall. Finally, Google Analytics also revealed that 82% of users accessed the site via a laptop or desktop computer, which has implications for web design.

Number of Ratings Performed

51 active users scored 4224 medial temporal lobes. Figure 9.8 presents a breakdown of the number of scores provided by each user by age range. 60% of the users scored more than half of the 160 medial temporal lobes in 80 images (left and right side rated separately). One person in each age band completed all ratings, while only a single user provided less than 20 ratings. On average, users took 17 seconds to score each MTL.

Feedback and Learning Points

After running the pilot study for four weeks, users were asked to complete a questionnaire about their experiences, 34 people responded. The survey findings and associated learning points are discussed below:

- **Motivations for taking part:** Over 50% of survey respondents were motivated

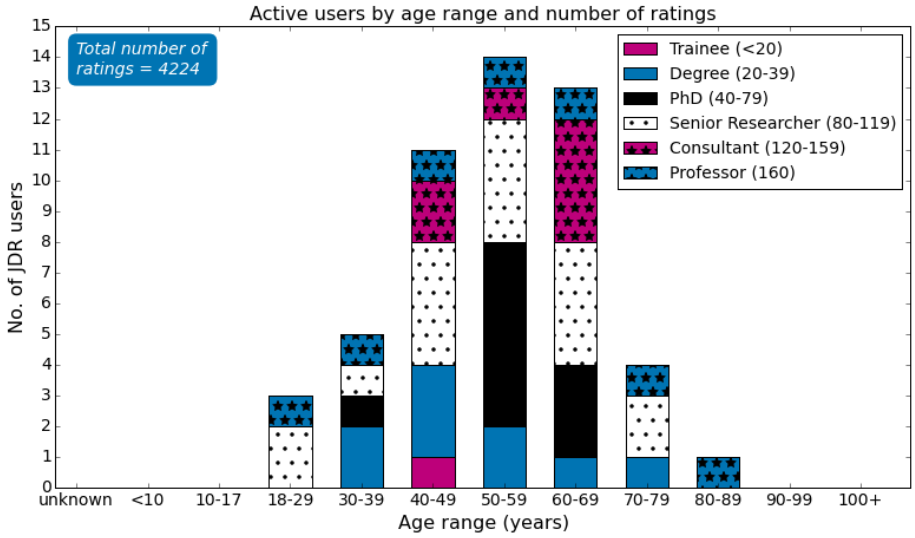


Figure 9.8: The number of medial temporal lobe ratings performed by the active users of Seeing Dementia. Within each age range the data is broken down by the number of ratings performed, e.g. trainees (plain pink bars) have provided less than 20 scores, professors (blue starred bars) have rated the left and right MTL in all images.

to take part in Seeing Dementia, and in dementia research more generally, by a family member developing dementia. Over 20% were motivated out of general interest and a desire to help. This has potentially important implications for the inclusion of informative content, as well as the design of future citizen science studies. Studies such as this one, with a more authentic scientific view of the data that is popular among the majority of citizen scientists, may be less appealing to an audience with a personal connection to the subject matter.

- **Understanding of the project:** Only 65% of respondents felt they fully understood the aims of the project, with 24% undecided, and 11% failed to understand the motivation. In terms of understanding the task instructions, 62% felt they understood the task, with 15% undecided, and 23% unable to understand what was required. Google Analytics revealed a disproportionately low number of views of the videos (introductory and tutorial) and informative content, suggesting that in the majority of cases users went straight into the task without instruction. Although there is a help box on the main rating page this is intended as a reminder and is probably insufficient in terms of instruction. This suggests that the in-

troduction of a mandatory training session before accessing the main rating task may be necessary to ensure users are aware of the task requirements.

- **Particular issues to be addressed:** Based on feedback in the survey, and emails received while the study was open, there are several usability issues to resolve. Firstly, the high proportion of people who registered but then failed to contribute any ratings needs to be addressed. The application was designed not to log people in automatically after registering, allowing them time to read the instructional materials before launching straight into the task. However, this confused many people and caused them to try and re-register. This was unforeseen and the underlying system logic produced unhelpful error messages, adding to the confusion. It is thought this was the primary reason for the high proportion inactive registered users. Secondly, the high proportion of users who were unsure or failed to understand the visual rating task is also a major issue. This response, coupled with the low number of views of the videos and informative content, suggests that the task itself should be presented more clearly, with the a compulsory training exercise built in. Survey respondents also suggested a preference for written guidance over video. Additional improvements to usability could be made by providing the user more regular feedback about how their performance compares to visual rating experts, or providing more information about the brain displayed in the image, and the disease processes affecting it. Despite the issues identified, the survey revealed 70% of respondents would recommend the project to a friend.

9.3.5 Future Work

This initial pilot study has revealed several interesting findings and opportunities to improve the Seeing Dementia application, as well as insights into the citizen science approach more generally. Many of the assumptions made during the design process appear to be incorrect, such as proportionately more tablet users than desktop users, and a preference for video rather than text instructions. This perhaps reflects a misunderstanding of an older user demographic, however, in such a small sample size it is uncertain how well these findings generalise to the national level. The issues outlined

above will all be addressed and retested using the newly formed JDR lay champions group.

More fundamental questions exist about the opportunities and motivations for applying citizen science in dementia research more generally. While the benefits in terms of raising awareness of dementia are unquestionable, it is difficult to strike a balance between answering valid scientific questions and providing a satisfying user experience. Traditional models of citizen science may be unsuitable in this context, not least because the user base is potentially older, with a personal connection to the subject matter. Prof. Lintott of Zooniverse advised that citizen science based on medical applications tends to be less popular as users often feel burdened by a responsibility to pick the "right" answer, which diminishes the user experience. Conversely "gamification" of the task is very expensive and without enough evidence to guarantee a successful outcome.

Continuing advances in mobile technologies may provide additional opportunities to better adapt citizen science to dementia research, allowing users to generate novel data to quantify the effect of dementia on the activities of daily life, or engaging patients and carers to build up an evidence base for low risk therapeutic interventions, e.g. approaches to music therapy.

9.4 Conclusions

Patients and carers take time out of their heavily burdened lives to provide researchers with data in the hope and belief that they will use this data responsibly to help, if not them, then future generations like them. It is vital to treat each donation with the respect it deserves and maximise the value that can be derived from it by accurately recording and sharing the data with the wider scientific community. In an attempt to meet these responsibilities, a scalable online data repository has been developed to store and share the data associated with the unique cohort collected in relation to this thesis. Further development of the application into a teaching and training tool may enhance the value that can be derived from this data that was generously donated by patients and their families. Citizen science offers unique opportunities to engage the general public in scientific research and raise awareness about dementia. The rapid development of mobile technologies has the potential to stimulate interest in this type

of approach. However, designing a successful citizen science project is non-trivial and much work is needed to better understand how this can be achieved.

Chapter 10

Thesis Conclusions

10.1 Introduction

Distinguishing between the different neurodegenerative causes of dementia is vitally important to allow affected individuals and their families to access appropriate treatment, support and care. This requirement will become even more pressing as/when disease-modifying therapies become available. While pathological biomarkers continue to evolve, in many cases their complexity and cost prohibit them from being implemented in routine clinical practice. It is therefore, important to optimising the use of the diagnostic techniques that are more widely available, such as MRI. The purpose of this thesis was to establish a large ante mortem imaging dataset of pathologically proven dementias, and use it to investigate ways of maximising the diagnostic value that can be extracted from routine T1-weighted imaging.

10.2 Can evidence from the literature be reliably identified through visual inspection at the individual patient level?

While there is a large evidence base detailing patterns of cerebral volume loss associated with specific neurodegenerative dementias, this information is rarely applied in routine clinical assessment. The potential reasons for this are threefold: 1) the majority of research studies are based on imaging acquired in clinically diagnosed dementias and may not accurately reflect underlying pathology; 2) findings are typically reported at the group level and their reliability is uncertain at the individual patient level; 3) radiology expertise and thorough knowledge of the neuroimaging literature are required in order to identify many of these features. To address the last of these issues, a structured

algorithm was developed to help guide visual assessment based on key findings in the literature. This was then tested by 5 analysts in an attempt to address the first two issues. The findings from this study suggest that while characteristic patterns of atrophy associated with each pathology predominated at the group level, there was variation at the individual patient level. Consistency among analysts in identifying the predominant pattern of atrophy was at best fair, although there was some indication that this might improve with experience. While this algorithm provides a useful summary of the current evidence in the literature relating to patterns of cerebral atrophy in neurodegenerative dementia, and may encourage more detailed reporting of findings among radiology experts, it is largely unsuitable for clinicians with less radiology experience. A more targeted, quantitative approach, introducing specific anatomical landmarks, is needed to make this information more accessible to a wider clinical audience.

10.3 Do visual rating scales improve diagnostic accuracy?

The results of this study suggest that visual rating scales, specifically designed to assess focal cerebral atrophy in patients with cognitive impairment, provide a useful framework for the reliable identification and interpretation of imaging findings in the differential diagnosis of dementia. The six previously described rating scales tested in this study (covering frontal, temporal and parietal lobe regions) were shown to be fairly reliable, but with some evidence, both in this study and in other published work, that this could be improved with training and experience. The average time taken to apply all six scales was less than three minutes. Using voxel based morphometry, the scores from each scale were shown to be highly correlated with the specific brain regions they were designed to assess, providing support for their regional specificity. Using the scales individually limited their diagnostic accuracy, however, combining these in an automated support vector classifier provided diagnostic accuracy equivalent to, and in some cases better than, unstructured visual assessment by dementia experts. These results suggests that visual rating scales offer clinicians without neuroradiology expertise a means of extracting diagnostically useful information in a time-efficient and inexpensive way, that is well suited for integration into routine clinical practice.

10.4 What patterns of regional atrophy are associated with specific molecular pathologies?

Using the latest voxel based morphometry software, this study provides a comprehensive assessment of differences in grey matter volume between pathologically distinct groups. At the time of writing this is thought to be the largest study of its kind in terms of sample size. Using effect size maps, this analysis demonstrates that neurodegenerative diseases have a global effect on brain tissue, a fact that is often overlooked when focussing only on statistically significant differences, which many studies are underpowered to detect. Unusually for studies of this kind, it was also possible to look at statistical differences between some groups using a very strict significance threshold, which convincingly confirmed previous reports in the literature imaging literature, including differences in cerebral atrophy between early and late onset Alzheimer's disease, and the relative lack of cerebral atrophy associated with DLB compared to other neurodegenerative pathologies. The patterns of atrophy identified here in ante mortem imaging of pathologically proven cases, could be used to inform future visual assessment studies, or automated region of interest analysis.

10.5 How accurate is fully automated diagnosis of dementia pathology?

Reliable fully automated image assessment could offer significant advantages in many isolated or under resourced healthcare settings. Using a combination of state of the art segmentation algorithms and support vector classifiers this study investigated the current diagnostic capabilities of this approach. Although the results of this analysis demonstrated reasonable diagnostic accuracy, there are several opportunities to improve on this performance which should be explored. In terms of the multi-atlas segmentation propagation, a template library more representative of the images under study may significantly improve segmentation accuracy. Questions also remain over the optimal level of detail to include in the atlas templates, with these results suggesting a reduction in classifier performance, both if the regions are larger and more generalised

or smaller and more fine grained. As previously noted by the Mayo clinic, which are thought to be the only other group with a comparably large imaging and pathology dataset, forming larger training and testing sets with more data would also significantly improve the accuracy of this approach and the reliability of analysis.

10.6 Future Work

The work of this thesis has produced a large imaging dataset of pathologically proven cases, a means of sharing this data with the wider scientific community, and initial analysis of the data to investigate opportunities to maximise the use of standard 3D T1-weighted imaging. There is however, considerable potential to expand the scope of this work. These opportunities are outlined below:

Pathology: From a practical perspective this study used the primary pathological diagnosis from histopathological assessment of brain tissue based on criteria in use at that time. The investigation of co-pathologies is limited not only by sample size, but also by potential short comings in historical assessment (compared to modern techniques) or the level of detail included in older pathology reports. Whilst the work of this thesis provides a framework to continue to increase the sample size, it would be of considerable interest to reclassify the older samples based on modern histopathological stains and consensus criteria.

Expanding the cohort: The data collated as part of this thesis provides a sizeable baseline for future study, while the web based data repository provides a framework to encourage greater collaboration and expansion of this collection to include prospective collection via clinical services in the sites already involved. This thesis has focused on T1 volumetric imaging but collection of other sequences and/or modalities could allow for similar approaches to explore, for example, the detailed pathology of vascular disease.

Incorporating clinical information: As previously stated, imaging should only be interpreted in the context of detailed clinical assessment. Whilst the clinical information collected to date is relatively minimal, it would be interesting to see if including this information in the classifier analysis could improve the utility of automated diagnosis.

The collection of prospective datasets and the use of age-matched control groups could help to further these aims.

Visual Rating: As demonstrated in Chapter 6, visual rating scales provide a framework for quickly extracting diagnostically useful information form standard structural imaging. It would be interesting to extend this analysis to investigate the diagnostic accuracy of visual rating scales in naive raters, and the subsequent effects of training. The online application described in Chapter 9 could also be adapted for use as an online visual rating training tool, providing initial/refresher training, and certification; with trainees validated not only against pathological diagnosis but also gold standard ratings by experts. From a technical perspective, it would also be worth investigating if initial registration to standard space improved inter-rater reliability. The decision was taken not to do this for this initial analysis, in an attempt to better match the research protocol to clinical conditions, however, the retrospective nature of this dataset, acquired over a twenty year period at multiple different centres, means there is probably far more variability in this dataset than is typically seen in one clinic. This may negatively influence both inter-rater reliability and scoring accuracy.

Novel regions of interest: The VBM and volumetric analysis described in Chapter 7 and Chapter 8 identified some potentially new findings of interest, particularly in relation to the motor areas differentiating DLB and Alzheimer's disease, and the brainstem and cerebellum which were neglected in our initial analysis. To what extent drawing attention to these regions may provide value on an individual patient basis is an interesting topic to explore further.

Technical advances: There are numerous technical advances that could be trialled such as developing more robust quality control procedures to accompany the multi-atlas segmentation propagation technique, and moving from binary to multi-level classifiers to enhance their utility as an automated diagnostic tool. Additionally, investigation into applying machine learning algorithms at the voxel level may result in a more efficient analysis pipeline and improvements in classification accuracy. Incorporating techniques to deal with vascular lesions (such as a non-local lesion filling strategy) prior to segmentation may also help to improve the accuracy of volumetric analysis. While efforts were made to exclude images demonstrating gross vascular pathology based on

visual inspection, smaller lesions may have had some influence of the results presented in Chapter 7 and Chapter 8.

10.7 Summary

This project has resulted in the development of the largest pathologically proven dementia-imaging cohort in Europe. Previously under-utilised retrospective data has been collated and cleaned, allowing this study and future studies to make better use of this valuable resource. Using a wide range of image analysis techniques, from simple visual assessment to more complex machine learning, various aspects of the dataset have been analysed with a view to maximising the diagnostic value of structural imaging routinely acquired in clinical assessment of dementia. A bespoke web application has also been developed to provide a means of sharing this dataset with the wider scientific community, both as a repository for post mortem imaging data, and as a radiology teaching resource.

Publications

The publications based on the results of this thesis, or that I have contributed to during my PhD are listed below.

Peer-Reviewed Journal Articles

- Harper, L., Barkhof, F., Scheltens, P., Schott, J. M. & Fox, N. C. An algorithmic approach to structural imaging in dementia. J Neurol Neurosurg Psychiatry 85, 692698 (June 2014).
- Slattery, C. F., Beck J. A., Harper, L et al. R47H TREM2 variant increases risk of typical early-onset Alzheimer's disease but not of prion or frontotemporal dementia. Alzheimers & Dementia 2014 Nov;10(6):602-608.e4 (Nov 2014).
- Harper, L., Barkhof, F., Fox, N. C. & Schott, J. M. Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales. Journal of Neurology, Neurosurgery, and Psychiatry (Apr. 2015).
- Harper, L., Fumagalli, G. G., Barkhof, F., Scheltens, P., O'Brien, J., Burton, E. J., Bouwman, F. H., Rohrer, J. D., Fox, N. C., Ridgway, G. R., & Schott, J. M. Maximising the value of structural MRI in the diagnosis of dementia: evaluation of visual rating scales in pathology proven cases. Brain (Accepted Dec 2015)

Book Chapters

- Harper, L. & Rossor, M. N. OECD Health Policy Studies Addressing Dementia: The OECD Response, Chapter 4 - The role of big data in driving global cooperation and innovation in dementia. OECD Publishing, (Mar 2015)

Conference Papers

- Harper, L., Burton, E. J., Bouwman, F. H., Barkhof, F., Scheltens, P., O'Brien, J., Fox, N. C., Ridgway, G. R., & Schott, J. M. Classification of pathology using brain substructure volumes in post mortem confirmed dementias. Alzheimer's &

Dementia: The Journal of the Alzheimer's Association , Volume 10 , Issue 4 , P32 - P33 (July 2014).

- Fumagalli, G. G., Harper, L., Gordon, E., Lehmann, M., Hyare, H., Warren, J. D., & Rohrer, J. D. Development of a visual rating scale for atrophy of the anterior cingulate, insula and frontal lobes. Am J Neurodegener Dis 3, 1375 (Oct 2014).

- Harper, L., Barkhof, F., Scheltens, P., O'Brien, J., Burton, E. J., Bouwman, F. H., Rohrer, J. D., Fox, N. C., Ridgway, G. R., & Schott, J. M. Distinct patterns of atrophy in post mortem confirmed dementias. Alzheimer's Association International Conference - oral presentation (July 2015)

- Harper, L., Fumagalli, G. G., Barkhof, F., Scheltens, P., O'Brien, J., Burton, E. J., Bouwman, F. H., Rohrer, J. D., Fox, N. C., Ridgway, G. R., & Schott, J. M. Visual assessment in post mortem proven dementias: clinical expertise versus visual rating. Alzheimer's Association International Conference - oral presentation (July 2015)

- Weston, P., Nicholas, J., Harper, L. et al. A cortical signature of familial Alzheimer's disease: cross-sectional and longitudinal study of presymptomatic changes. Alzheimer's Association International Conference - oral presentation (July 2015)

- Weston, P., Nicholas, J., Harper, L. et al.Measurement of cortical mean diffusivity detects early microstructural breakdown of the cerebral cortex in presymptomatic familial Alzheimer's disease. Alzheimer's Association International Conference - oral presentation (July 2015)

Appendix A

Six Isoforms of Tau

Microtubules, one of three principle components of the cytoskeleton, determine cell shape and a variety of cell movements [318]. Normally functioning tau proteins regulate microtubule assembly in brain cells, influencing their rigidity, length, stability and ability to interact with other organelles [18]. There are six functionally similar tau proteins (isoforms) in the brain that differ by the presence of either three or four repeat regions in the carboxy-terminal part of of the molecule and the presence or absence of one or two inserts in the amino terminal part (see Figure A.1) [18]. Four-repeat (4R) isoforms are more efficient at promoting microtubule assembly [18], although 4R and three-repeat (3R) isoforms are typically expressed in a one-to-one ratio in the adult brain (only the shortest tau isoform (0N3R) is found in the foetal brain) [319]. Microtubule assembly is also influenced by phosphorylation state, with phosphorylated tau proteins less efficient than their non-phosphorylated counterparts.

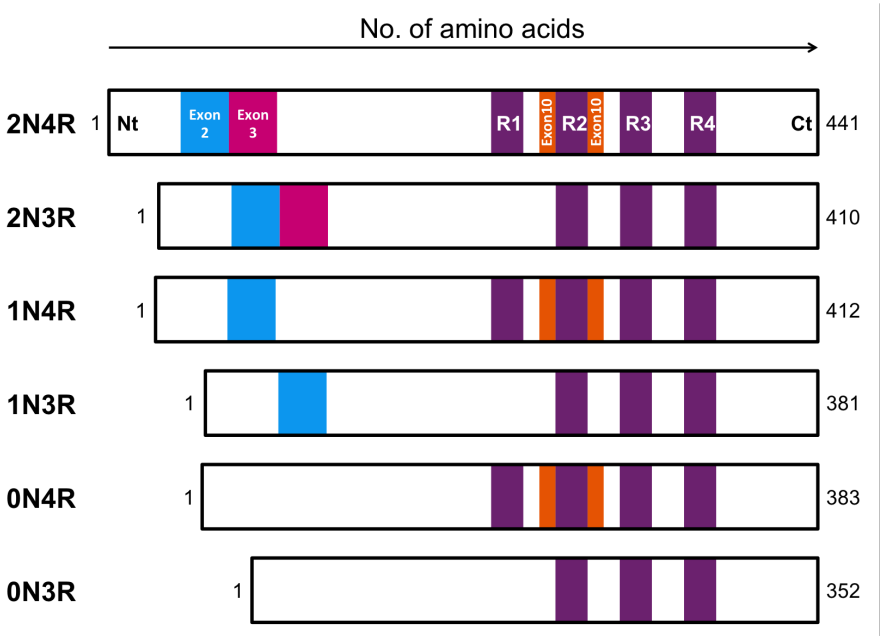


Figure A.1: Six isoforms of tau present in the developed brain. Only the shortest isoform (0N3R) is present in the foetal brain. Nt = Amino (N) terminal, Ct = Carboxyl (C) terminal, R1-R4 = repeat regions

Appendix B

Visual Rating Design, Methodology and Validation

There are a number of factors implicit in the design of visual rating scales that may determine their successful adoption in within the research community, and their potential for adoption into clinical practice. These factors are discussed below and summarised in Table B.1.

Defining and displaying ROIs

The brain regions selected for visual rating have the greatest impact on the usefulness of the scale. Regions should be selected based on established findings from volumetric image analysis and/or macroscopic pathological assessment of the disease population of interest. The number of regions to be rated, the number of imaging planes to assess (axial, sagittal or coronal) and the number of slices used is likely to impact on the reliability of the scale, with reliability decreasing with increasing scale complexity. Specifying landmark identifiable slices for rating helps to ensure consistency between raters. There is good rationale for including focal regions, such as the MTL, which are typically preferentially involved in certain conditions, for example, Alzheimer's disease, and have been shown to correlate with clinical measures of disease severity such as mini mental state examination (MMSE) [129]. Choice of MR pulse sequence affects both the appearance of atrophy and the visible extent of white matter changes and should also be specified. T1-weighted images offer good grey-white matter and CSF contrast, with high resolution three-dimensional volume acquisitions (that can be reconstructed in all three planes) offering the greatest utility for rating atrophy. T2-weighted images are less reliable, since the amount of CSF can be overestimated if T2-weighting is too strong. Image quality will also affect the reliability of the scale, with rating less reliable on scans that are subject to artefacts. Consistent image slice positioning will also help

Design decisions	Methodological considerations	Clinical implications
Defining a region of interest	Is there a good evidence base for atrophy in the region?	Dictates utility and interpretation in certain clinical populations.
	Should the region be rated in >1 imaging plane?	Requires three-dimensional or multiple image acquisitions and increases time to perform rating.
	Is there an imaging landmark to allow consistent slice selection?	Improves test-retest reliability
Displaying a region of interest	Is the MR contrast appropriate and in common clinical use?	Affects the appearance of atrophy and the sensitivity to artefacts
	Is the appearance of the region badly affected by patient positioning?	Difficult to rate certain regions or to reliably assess symmetry if the head is tilted
Defining scale increments	How much variation can reliably be captured?	Truncated use of the scale may result in decreased diagnostic value
	Is there a reliable cut-off between normal and abnormal scan appearance?	Affects clinical interpretation
	Should the cut-off be adjusted for age?	
Providing training material	How is each scale increment best described?	Provides a useful framework for scoring
	Are reference images available?	
	Are there expert raters available to provide training sets?	Provides confidence in ratings and a means of audit
Validating the scale	What is the inter-rater/intra-rater reliability and how should it be measured?	Determines suitability for use in clinical practice and comparison with other scores
	Do the scores correlate with clinical measures or other measures of atrophy?	Validates clinical relevance
	Is there a diagnostic gold standard available for comparison?	Provides validation of diagnostic value

Table B.1: Summary of key design decisions, associated methodological considerations and clinical implications associated with visual rating scales

to improve the reliability of the scale.

Scale increments

The number of scale increments influences the level of detail captured by the scale. A balance must be struck between detailed quantification and the degree of change that can be reliably differentiated by visual inspection. In terms of structural neuroimaging, a four-point or five-point scale is most commonly used. The scale is typically dichotomised to classify normal and abnormal scan appearance. In both four-point and five-point scales, scale points 0 and 1 typically represent the degree of variation within the normal population, with points 2 and above describing more obvious pathological change. Four-point scales force the rater to make a more definite choice of disease state (presence or absence), therefore, increasing specificity at the expense of sensitivity. Five-point scales on the other hand may be more sensitive to earlier stages of disease but may also increase the number of false-positive results. In terms of the scales developed for use in the diagnosis of dementia, five-point scales may be particularly sensitive to the effects of ageing. Using age-specific cut-offs may help to improve scale accuracy [136, 137, 320].

The effect of training

Training can have a significant affect on the performance of the scale. Reference images providing examples of each scale point are particularly useful and are likely to impact positively on the reliability of the scale. Reference images which include delineation of ROIs, such as those provided by Urs et al could also help to improve reliability, particularly among less experienced raters or raters without radiology expertise. Detailed descriptions of the expected appearance for each point on the scale can also be helpful to guide raters and improve consistency. Training sets representative of the clinical or study population, pre-rated by expert raters, would help to ensure high observer agreement before implementation into clinical practice or research protocols. Training sets can also be used to audit rater reliability at defined intervals or after a period of absence [115].

Validation

If rating scales are used as a method of measurement to make inferences about disease state, it is important that both the measurement technique and validation of the technique is rigorous. Test-retest studies are essential to determine the (inter-rater/intra-rater) reliability of the scale. Appropriate statistical procedures should be applied and fully reported to allow clear interpretation of the results and fair comparison with other studies. However, if used routinely, the affect of training and rater experience is likely to improve the reliability of the scale. Correlation with clinical measures of cognition [126, 129, 321] and volumetric measurements [127, 145, 292, 322] are also useful to help validate the scale. Diagnostic tests should also be validated against an established gold standard measurement technique. Currently, with the exception of individuals with genetic mutations, post mortem examination of brain tissue is the only definitive means of establishing diagnosis in neurodegenerative dementia. In most scales described here, classification of disease groups, and therefore measures of scale sensitivity and specificity, are based on clinical diagnosis of the study population.

Appendix C

Visual Assessment Form

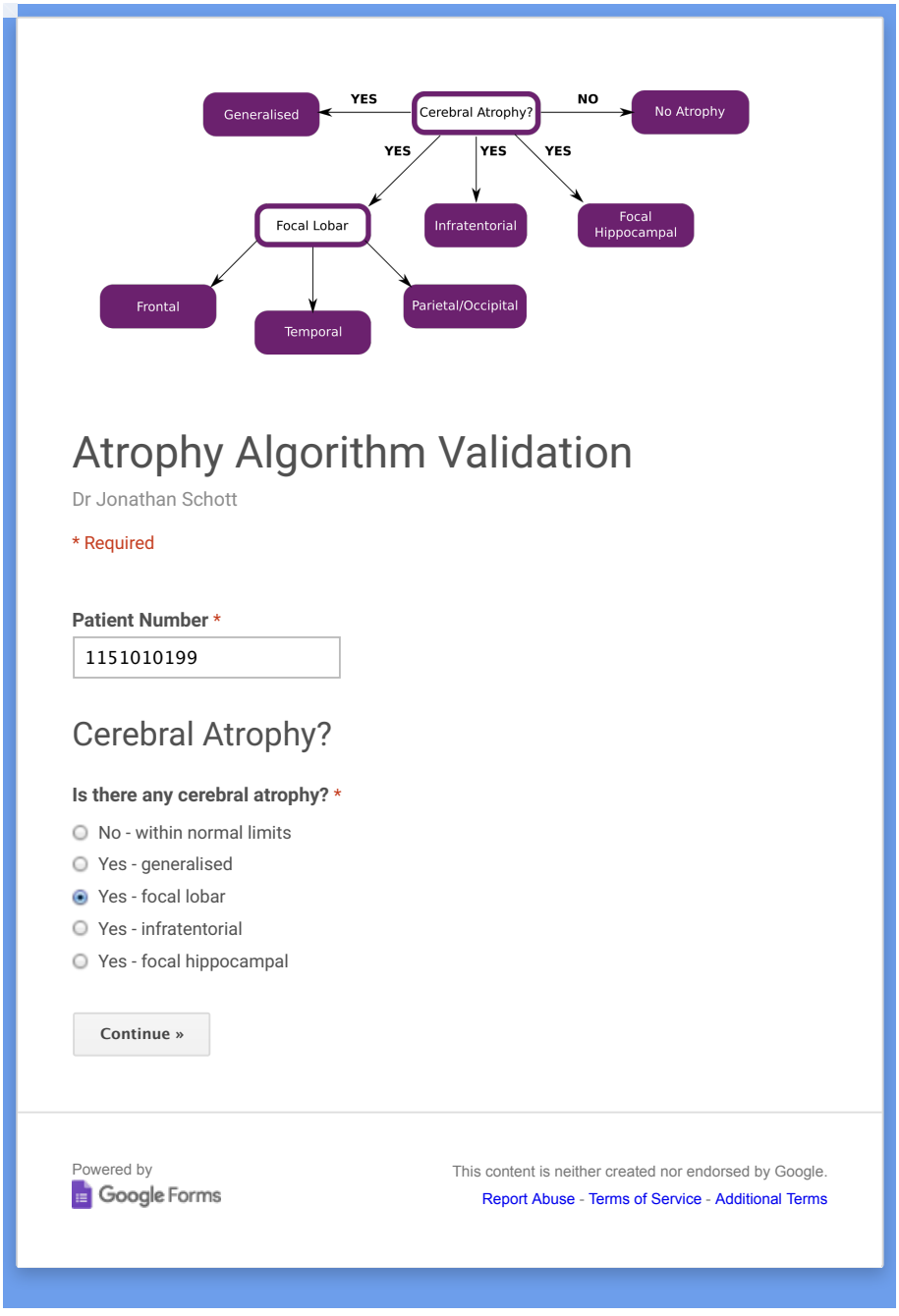


Figure C.1: First page of the visual assessment form. Patient ID is manually populated, in sync with the MIDAS image viewer

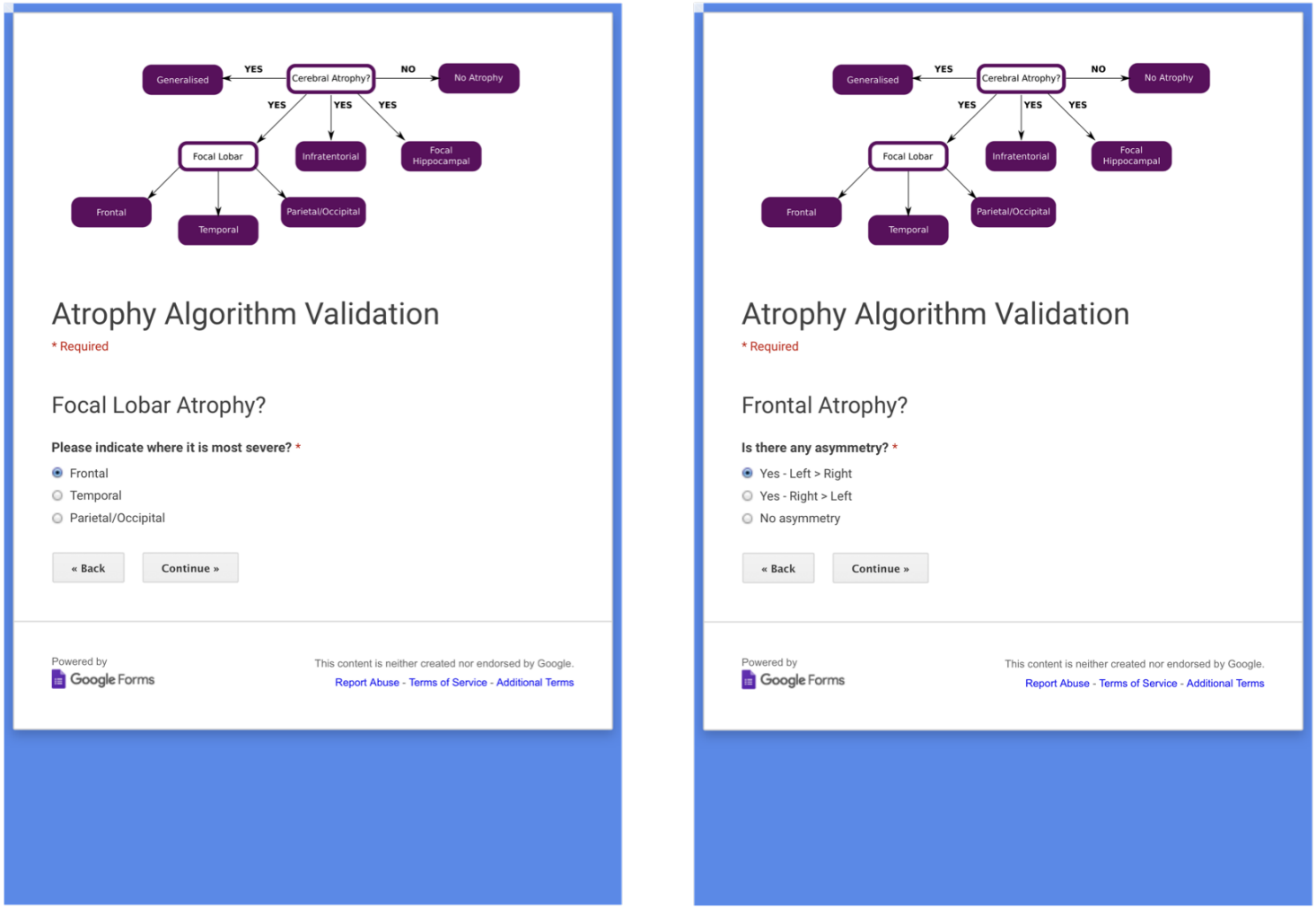


Figure C.2: The form progresses based on the answers provided, in accordance with the algorithm.

Appendix D

Fleiss' Kappa Calculation and 95% Confidence Intervals

As discussed in Chapter 5, kappa statistics are defined as

$$\kappa = \frac{Pa - Pe}{1 - Pe}$$

Where Pa is the proportion of observations in agreement between analysts and Pe is the proportion in agreement due to chance. In terms of Fleiss' kappa, for every scan i=1,..,n and every category j=1,..,k, X_{ij} = the number of analysts that assign category j to scan i, and Pa is defined as

$$Pa = \frac{1}{mn(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} X_{ij}^2 - mn$$

Pe is given by

$$Pe = \sum_{j=1}^{k} q_j^2 \qquad where \qquad q_j = \frac{1}{mn} \sum_{j=1}^{k} X_{ij}$$

Kappa values can also be calculated for every category by

$$\kappa_j = 1 - \frac{\sum_{i=1}^{k} X_{ij}(m - X_{ij})}{mn(m-1)q_j(1-q_j)}$$

The standard error for each κ_j is given by

$$se_j = \sqrt{\frac{2}{mn(m-1)}}$$

Then the standard error for κ is given by

$$se = se_j \frac{\sqrt{[\sum_{i=1}^{k} q_j(1-q_j)]^2 - \sum_{i=1}^{k} q_j(1-q_j)(1-2q_j)}}{\sum_{i=1}^{k} q_j(1-q_j)}$$

The test statistic z = κ/se can be approximated by the standard normal distribution, allowing a confidence interval to be calculated. 95% confidence intervals were calculated in Python using the percent point function (ppf) (or the inverse cumulative distribution function).

$$95\%\ Confidence\ Interval = \kappa \pm se * norm.ppf(0.05/2)$$

Appendix E

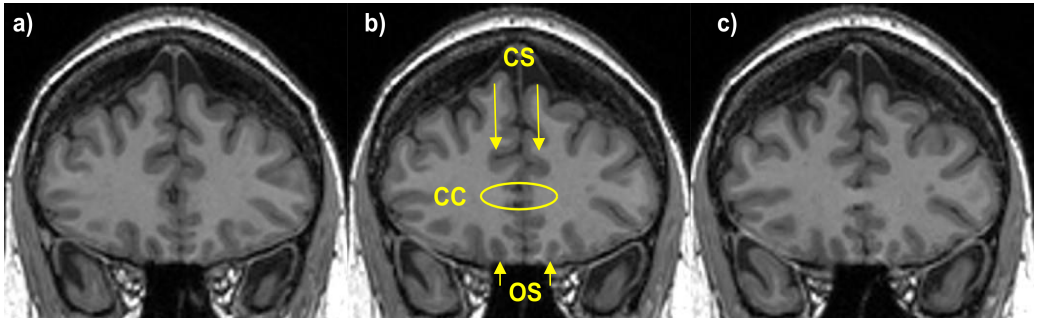
Visual Rating Protocol

Reference guides for the six visual rating scales were provided as follows.

Orbito-Frontal Rating Protocol

Slice Selection

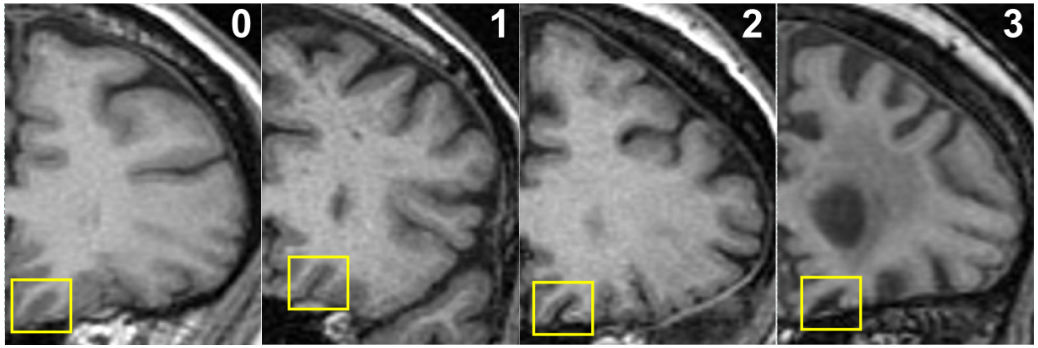
a) Corpus callosum not yet visible (pre-rating slice)
b) Corpus callosum just visible (rate olfactory sulcus and cingulate sulcus on this slice)
c) Post-rating slice



CC - corpus callosum, OS - olfactory sulcus, CS - cingulate sulcus

Rating Guide

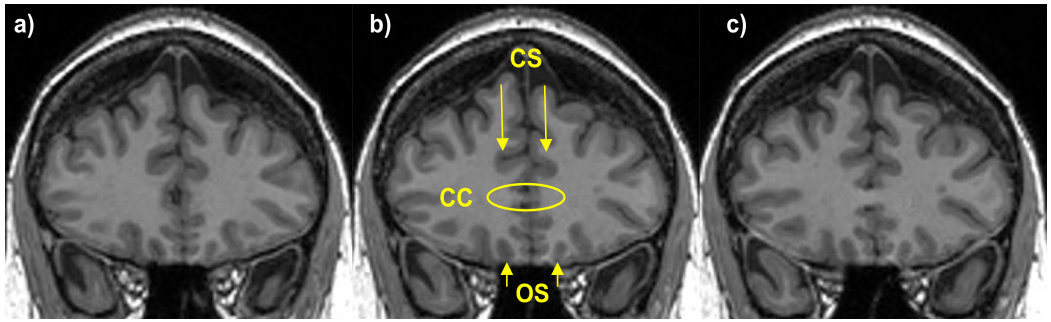
0: Closed sulcus
1: Small sulcal slit, just revealing CSF
2: Opening of the sulcus, CSF clearly visible
3: Severe widening of the sulcus



Rostral Anterior Cingulate Rating Protocol

Slice Selection

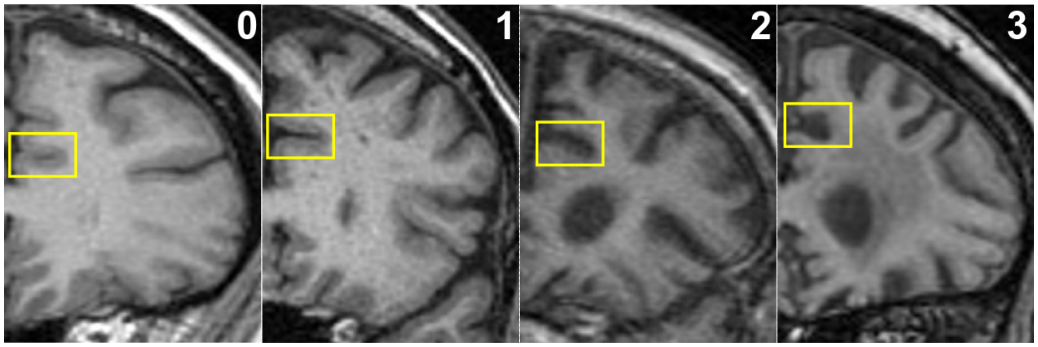
a) Corpus callosum not yet visible (pre-rating slice)
b) Corpus callosum just visible (rate olfactory sulcus and cingulate sulcus on this slice)
c) Post-rating slice



CC - corpus callosum, OS - olfactory sulcus, CS - cingulate sulcus

Rating Guide

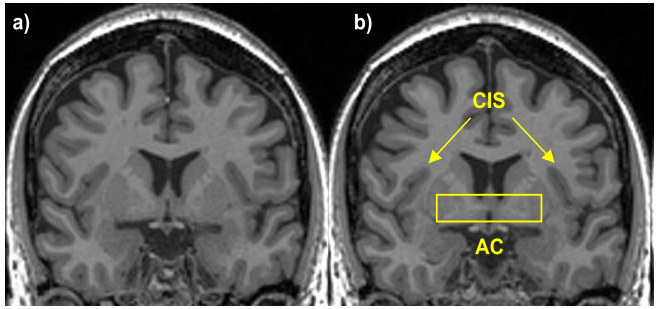
0: Closed sulcus
1: Sulcal opening (CSF visible), although narrower towards the peak
2: Sulcal widening along the length of the sulcus
3: Severe widening of the sulcus



Fronto-Insula Rating Protocol

Slice Selection

a) Anterior commissure (AC) not yet visible (pre-rating slice)
b) Anterior commisure just visible (rate this slice and the 2 posterior)

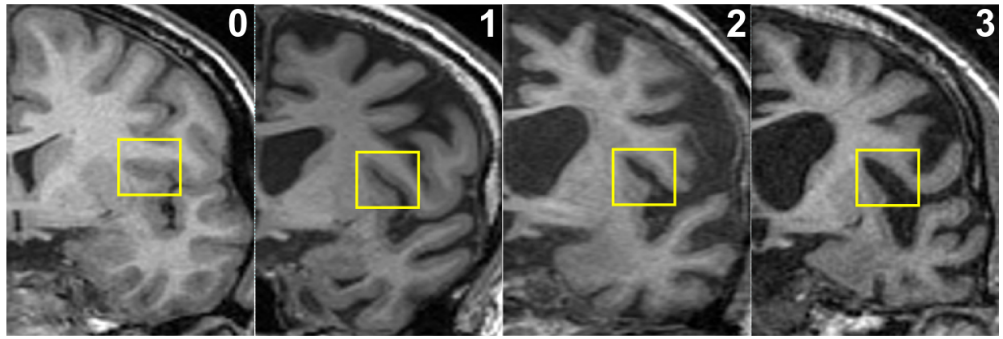


CIS: Circular insular sulcus, AC: Anterior commissure

Rating Guide

(Average the score over the 3 slices)

0: Closed sulcus
1: Sulcal opening, CSF clearly visible
2: Sulcal widening and the emergence of an arrow head shape pointing towards the midline
3: Severe widening along the length of the sulcus



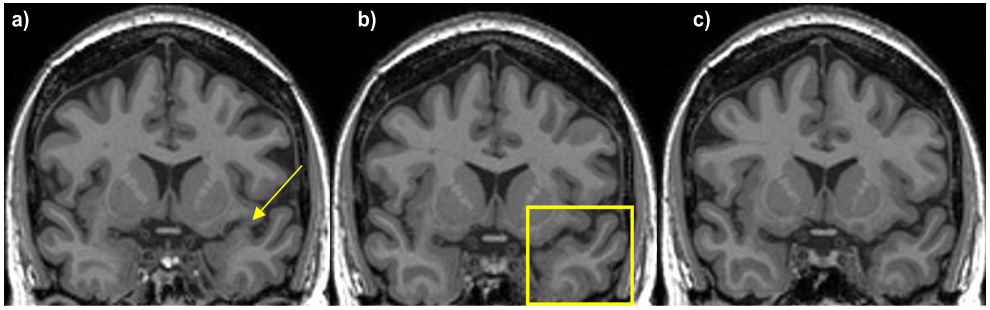
Anterior Temporal Rating Protocol

Slice Selection

a) Connection between the frontal and temporal lobes is still visible (pre-rating slice)

b) No visible connection between the frontal and temporal lobes (rate this slice)

c) Post-rating slice



Rating Guide

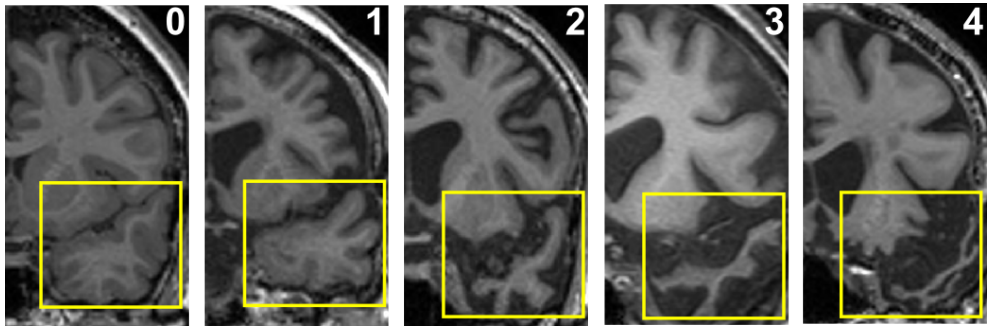
0: Normal appearances

1: Slight prominence of anterior temporal sulci

2: Temporal sulci definitely widened

3: Gyri severely atrophic and ribbon-like. WM and GM cannot be distinguished (normal temporal lobe at this level is less substantial than the frontal lobe, ribbon-like gyri of stage 3 temporal lobe are similar to stage 4 frontal gyri)

4: Temporal pole has a simple linear profile or is not seen at all



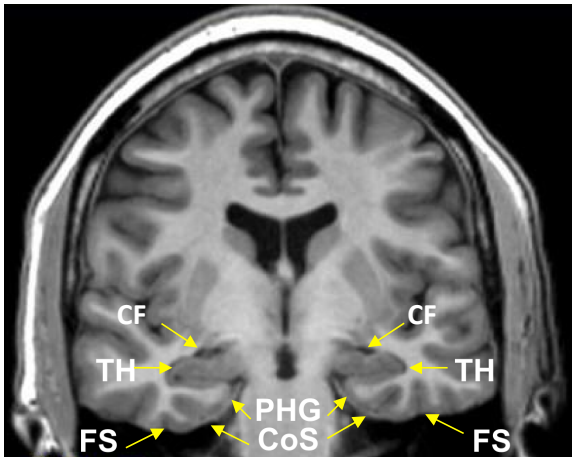
Medial Temporal Rating Protocol

Slice Selection

- **In the middle of the hippocampal body, in front of the pons or halfway through the pons depending on the angle of the scan**
- **Scroll though the hippocampus to get an impression of the atrophy throughout**
- **Don't rate too close to the amygdala. If the hippocampus curls up, the slice is too close to the hippocampal head.**
- **At the origin of the fornix, the slice is too close to the tail.**
- **A score of 0 can still be given if there is some opening of the choroid fissure on a few slices through the hippocampal body if the remainder are closed.**
- **A score of 1 is given if the choroid fissure is opened over the entire length of the hippocampal body.**

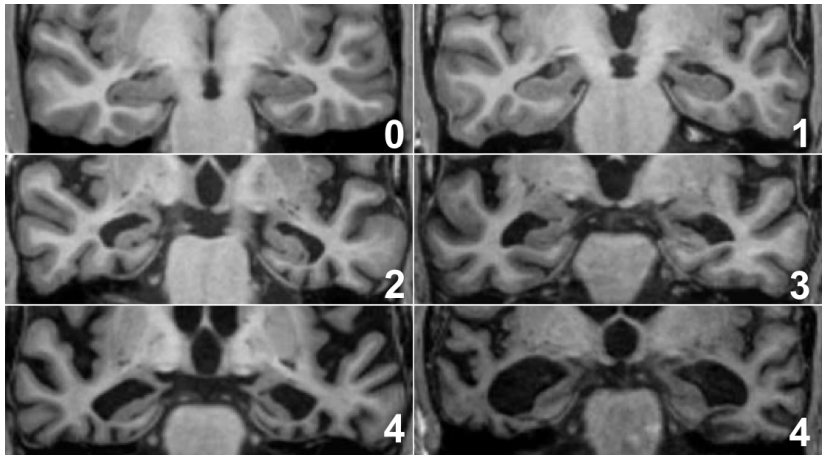
Rating Guide

(References images from http://www.radiologyassistant.nl/en/p43dbf6d16f98d/dementia-role-of-mri.html)



Also take these structures into account:

CF – choroid fissure
TH - temporal horn
PHG - parahippocampal gyrus
CoS - collateral sulcus
FS - fusiform sulcus



0: Normal,
1: Widened choroid fissure
2: Increased widening of the choroid fissure, widening of the temporal horn, opening of other sulci (i.e. collateral/fusiform sulcus)
3: Pronounced volume loss of the hippocampus,
4: End stage atrophy

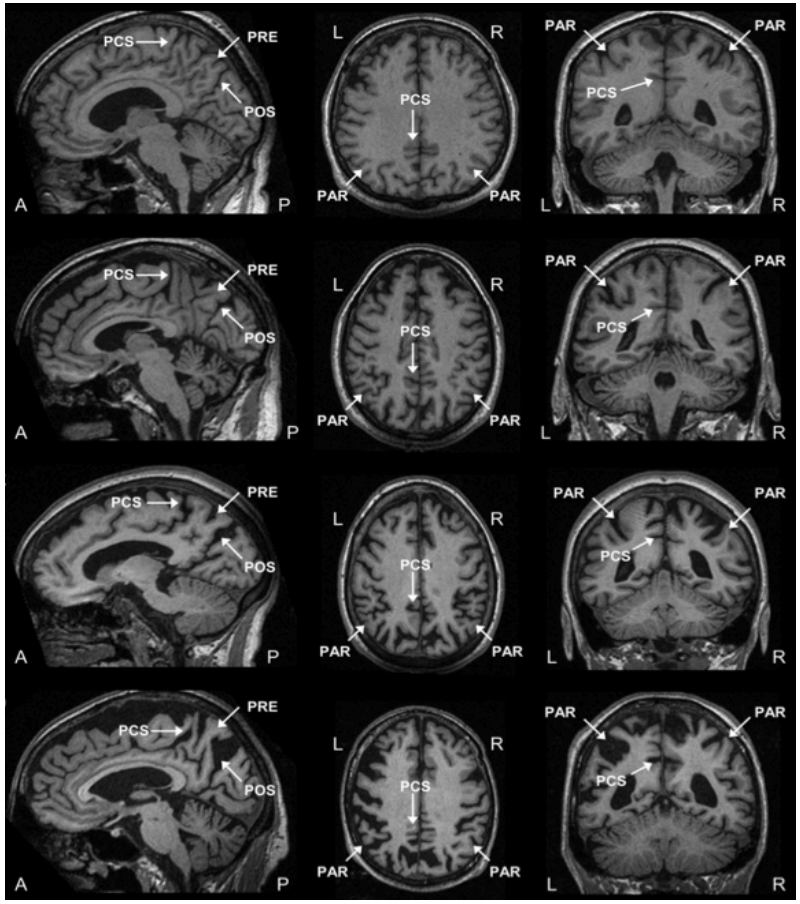
Posterior Atrophy Rating Protocol

0

1

2

3



Rating Guide

0: Closed sulci of parietal lobes and cuneus
1: Mild widening of posterior cingulate and parieto-occipital sulci
2: Substantial widening of the sulci
3: Extreme widening of the posterior cingulate and parieto-occipital sulci

Slice Selection

No slice selection – just scroll through
PAR - parietal lobe
PCS - posterior cingulate sulcus
POS - parieto-occipital sulcus
PRE -precuneus

Image from Lehmann et al, Neurobiol Aging. 2012 Mar;33(3):627.e1-627.e12.

Appendix F

Hammers' Atlas

Numbers2regions for the Hammers_mith atlas n30r83

The region list on the next page translates the (essentially arbitrary) voxel value coding for the regions into anatomical names.

The exact region definitions and protocols used for manually delineating the 30x83 regions can be found in the following papers:

(for regions 1-49):
Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* 2003, 19: 224-247.

(for regions 50-83):
Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, Edwards AD, **Hammers A**. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *Neuroimage* 2008 Apr 1;40(2):672-684
and on the associated web page, www.brain-development.org.

IMPORTANT HINT:
Odd numbers always indicate a region on the right side of the brain; even numbers one on the left. This can help when using n30r83 with new software, to check R/L orientation is correct after spatial manipulations.

Alexander Hammers, Lyon, 15 December 2011.

Number in Atlas	Name of Structure
Temporal Lobe	
1; 2	Hippocampus
3; 4	Amygdala
5; 6	Anterior temporal lobe, medial part
7; 8	Anterior temporal lobe, lateral part
9; 10	Parahippocampal and ambient gyri
11; 12	Superior temporal gyrus, posterior part
13; 14	Middle and inferior temporal gyrus
15; 16	Fusiform gyrus
30; 31	Posterior temporal lobe
82; 83	Superior temporal gyrus, anterior part
Posterior Fossa	
17; 18	Cerebellum
19	Brainstem
Insula and Cingulate gyri	
20; 21	Insula
24; 25	Cingulate gyrus (gyrus cinguli), anterior part
26; 27	Cingulate gyurs (gyrus cinguli), posterior part
Frontal Lobe	
28; 29	Middle frontal gyrus
50; 51	Precentral gyrus
52; 53	Straight gyrus
54; 55	Anterior orbital gyrus
56; 57	Inferior frontal gyrus
58; 59	Superior frontal gyrus
68; 69	Medial orbital gyrus
70; 71	Lateral orbital gyrus
72; 73	Posterior orbital gyrus
76; 77	Subgenual frontal cortex
78; 79	Subcallosal area
80; 81	Pre-subgenual frontal cortex
Occipital Lobe	
64; 65	Lingual gyrus
66; 67	Cuneus
22; 23	Lateral remainder of occipital lobe
Parietal Lobe	
60; 61	Postcentral gyrus
62; 63	Superior parietal gyrus
32; 33	Inferiolateral remainder of parietal lobe
Central Structures	
34; 35	Caudate nucleus
36; 37	Nucleus accumbens
38; 39	Putamen
40; 41	Thalamus
42; 43	Pallidum
44	Corpus callosum
74; 75	Substantia nigra
Ventricles	
45; 46	Lateral ventricle (excluding temporal horn)
47; 48	Lateral ventricle, temporal horn
49	Third ventricle

Appendix G

Neuromorphometrics/BrainCOLOR Atlas

Volume No.	Label
[4]	3rd Ventricle
[11]	4th Ventricle
[15]	5th Ventricle
[23]	Right Accumbens Area
[30]	Left Accumbens Area
[31]	Right Amygdala
[32]	Left Amygdala
[35]	Brain Stem
[36]	Right Caudate
[37]	Left Caudate
[38]	Right Cerebellum Exterior
[39]	Left Cerebellum Exterior
[40]	Right Cerebellum White Matter
[41]	Left Cerebellum White Matter
[42]	Right Cerebral Exterior
[43]	Left Cerebral Exterior
[44]	Right Cerebral White Matter
[45]	Left Cerebral White Matter
[46]	CSF
[47]	Right Hippocampus
[48]	Left Hippocampus
[49]	Right Inf Lat Vent
[50]	Left Inf Lat Vent
[51]	Right Lateral Ventricle
[52]	Left Lateral Ventricle
[55]	Right Pallidum
[56]	Left Pallidum
[57]	Right Putamen
[58]	Left Putamen
[59]	Right Thalamus Proper
[60]	Left Thalamus Proper
[61]	Right Ventral DC
[62]	Left Ventral DC
[63]	Right vessel
[64]	Left vessel
[65]	Left Insula
[66]	Right Insula
[69]	Optic Chiasm
[71]	Cerebellar Vermal Lobules I-V
[72]	Cerebellar Vermal Lobules VI-VII
[73]	Cerebellar Vermal Lobules VIII-X
[74]	Vitamin E Tablet
[75]	Left Basal Forebrain
[76]	Right Basal Forebrain
[100]	Right ACgG anterior cingulate gyrus

[101] Left ACgG anterior cingulate gyrus
[102] Right AIns anterior insula
[103] Left AIns anterior insula
[104] Right AOrG anterior orbital gyrus
[105] Left AOrG anterior orbital gyrus
[106] Right AnG angular gyrus
[107] Left AnG angular gyrus
[108] Right Calc calcarine cortex
[109] Left Calc calcarine cortex
[112] Right CO central operculum
[113] Left CO central operculum
[114] Right Cun cuneus
[115] Left Cun cuneus
[116] Right Ent entorhinal area
[117] Left Ent entorhinal area
[118] Right FO frontal operculum
[119] Left FO frontal operculum
[120] Right FRP frontal pole
[121] Left FRP frontal pole
[122] Right FuG fusiform gyrus
[123] Left FuG fusiform gyrus
[124] Right GRe gyrus rectus
[125] Left GRe gyrus rectus
[126] Right IFG inferior frontal gyrus
[127] Left IFG inferior frontal gyrus
[128] Right IOG inferior occipital gyrus
[129] Left IOG inferior occipital gyrus
[132] Right ITG inferior temporal gyrus
[133] Left ITG inferior temporal gyrus
[134] Right LiG lingual gyrus
[135] Left LiG lingual gyrus
[136] Right LOrG lateral orbital gyrus
[137] Left LOrG lateral orbital gyrus
[138] Right MCgG middle cingulate gyrus
[139] Left MCgG middle cingulate gyrus
[140] Right MFC medial frontal cortex
[141] Left MFC medial frontal cortex
[142] Right MFG middle frontal gyrus
[143] Left MFG middle frontal gyrus
[144] Right MOG middle occipital gyrus
[145] Left MOG middle occipital gyrus
[146] Right MOrG medial orbital gyrus
[147] Left MOrG medial orbital gyrus
[148] Right MPoG postcentral gyrus medial segment
[149] Left MPoG postcentral gyrus medial segment
[150] Right MPrG precentral gyrus medial segment

[151] Left MPrG precentral gyrus medial segment
[152] Right MSFG superior frontal gyrus medial segment
[153] Left MSFG superior frontal gyrus medial segment
[154] Right MTG middle temporal gyrus
[155] Left MTG middle temporal gyrus
[156] Right OCP occipital pole
[157] Left OCP occipital pole
[160] Right OFuG occipital fusiform gyrus
[161] Left OFuG occipital fusiform gyrus
[162] Right OpIFG opercular part of the inferior frontal gyrus
[163] Left OpIFG opercular part of the inferior frontal gyrus
[164] Right OrIFG orbital part of the inferior frontal gyrus
[165] Left OrIFG orbital part of the inferior frontal gyrus
[166] Right PCgG posterior cingulate gyrus
[167] Left PCgG posterior cingulate gyrus
[168] Right PCu precuneus
[169] Left PCu precuneus
[170] Right PHG parahippocampal gyrus
[171] Left PHG parahippocampal gyrus
[172] Right PIns posterior insula
[173] Left PIns posterior insula
[174] Right PO parietal operculum
[175] Left PO parietal operculum
[176] Right PoG postcentral gyrus
[177] Left PoG postcentral gyrus
[178] Right POrG posterior orbital gyrus
[179] Left POrG posterior orbital gyrus
[180] Right PP planum polare
[181] Left PP planum polare
[182] Right PrG precentral gyrus
[183] Left PrG precentral gyrus
[184] Right PT planum temporale
[185] Left PT planum temporale
[186] Right SCA subcallosal area
[187] Left SCA subcallosal area
[190] Right SFG superior frontal gyrus
[191] Left SFG superior frontal gyrus
[192] Right SMC supplementary motor cortex
[193] Left SMC supplementary motor cortex
[194] Right SMG supramarginal gyrus
[195] Left SMG supramarginal gyrus
[196] Right SOG superior occipital gyrus
[197] Left SOG superior occipital gyrus
[198] Right SPL superior parietal lobule
[199] Left SPL superior parietal lobule
[200] Right STG superior temporal gyrus

[201] Left STG superior temporal gyrus
[202] Right TMP temporal pole
[203] Left TMP temporal pole
[204] Right TrIFG triangular part of the inferior frontal gyrus
[205] Left TrIFG triangular part of the inferior frontal gyrus
[206] Right TTG transverse temporal gyrus
[207] Left TTG transverse temporal gyrus

Appendix H

Open-Data

The following is an extract from OECD Health Policy Studies Addressing Dementia: The OECD Response, Chapter 4 - The role of big data in driving global cooperation and innovation in dementia. OECD Publishing, (Mar 2015)

The Internet has not only transformed expectations about knowledge transfer and data sharing but also the timescales on which they occur. Rapid access to shared resources has led to extraordinary growth and development in many fields from basic science to big business, however, with the exception of the great accomplishments in genetics, medical research has largely failed to capitalise on these opportunities. Instead, the field has stuck closely to traditional models of research where unique data is the primary commodity available to generate funding and subsequently publications, and the ability to sustain this circular economy ultimately determines the success of most labs. To move forward, it is vital to examine these cultural barriers and look for inspiration from other fields on how to begin to dismantle them.

At the outset of the Human Genome Project (HGP), realising the societal significance of the data that would be generated, the genetics community disrupted traditional models of scientific practice by agreeing on a ground-breaking set of principles [323]. The Bermuda Principles, as they became known, called for all human genomic sequence information to be freely available in the public domain and rapidly released, in some cases automatically and within 24 hours. These principles have subsequently been reaffirmed and extended to include data from other large scale community resource projects in genomics, transcriptomics, proteomics and metabolomics [324–326]. Enabling such a radical change in practice required compliance from all stakeholders, establishing clear responsibilities for resource producers, resource users, funding agencies and publication streams. Omics research has recognised the profound value of making the vast amounts of data generated quickly and widely available to scientists, to achieve results beyond what the data producers themselves could produce within the same time pe-

riod, and often beyond the scope of the original project. Conversely, there is increasing awareness in dementia research that it is impossible to generate the wealth of data required to understand the complexities of neurodegeneration without sharing resources [327], at all levels of investigation, from genomics to clinical research. Moreover, if we are to discover disease-modifying therapies before dementia overwhelms the resources for care available in our ageing society, we must do this quickly, agreeing on a similarly bold set of principles to overcome the disincentives to share that are inherent in current research practices. To emulate the Omics model, the dementia research community must first identify data sources suitable for rapid release that would best serve the community. A useful starting point may be a publication portal for negative results. Negative results are seldom disseminated despite having the potential to significantly reduce duplication of effort, make better use of valuable resources, and, as a result, accelerate scientific discovery [328]. The provision of a searchable database and a straightforward publication template could help to re-evaluate current models of scientific communication. Additional guidelines for use in terms of both data entry, and due diligence searching as part of a grant application process, could help ensure such a provision was used to its full potential.

More challenging models of early access are based on the pre-publication release of data. The benefits of this approach may be far reaching, not only increasing the rate of scientific discovery, but also helping to address one of the major problems in science; the failure to replicate results [329]. Rapid publication of methods and data could allow results to be confirmed or refuted by other groups, preventing flawed methods being carried through to final publication, and producing greater confidence in those that are. Not only could this idea have a major impact on drug discovery, it would also provide a novel and more responsible publication route, potentially relieving researchers from the current pressure to publish quickly by offering greater recognition for due diligence.

In addition, pre-publication could also extend to cohort data from longitudinal studies. Data release could be scheduled after every timepoint or released in batches based on acquisition of an agreed number of participants. By encouraging multi-centre collaboration and harmonisation of data collection tools this model could be extended to

include many more data points, contributing to a larger more useful pool of data and essentially creating new, larger cohorts than have previously been assembled. Moreover, as dementia research moves towards preclinical trials there is an opportunity to build on work already started in population based studies and encourage greater collaboration and integration with social sciences and epidemiological research groups who already have well-established protocols for data sharing [330].

To move towards greater sharing of resources and faster paced development it would be prudent to also consider the benefits of sharing analysis tools, software and computing resources. In the case of wet biomarker data, such as CSF, where lab based analysis can have a major impact on the reported measure [331], it may be useful to set up centres of excellence conforming to agreed lab standards, where locally acquired samples can be sent for analysis and the results subsequently made available to the wider community. In terms of computational analysis, state of the art algorithms and tools could be made more widely available by pooling valuable computational resources and harnessing the power of cloud computing.

As the Omics community have demonstrated, successful and timely data sharing hinges on a system of quadripartite responsibilities between resource producers, resource users, funding bodies and publishers. Incentives for change must primarily safeguard the interests of data generators whilst ensuring the economic benefits of shared resources and the increased pace of discovery are experienced by the entire research community and, more importantly, by society as a whole. Reasonably straightforward options include introducing unique publication opportunities around negative results, the detailed description of the data generation, and the development of best practice guidelines for data collection and harmonisation of the collection tools currently available. On a technical level, investment is needed to ensure existing data sharing frameworks in both commercial and publically funded research are useful, efficient, fully integrated and suitable for sustained use by the entire research community.

Patients and carers take time out of their heavily burdened lives to provide researchers with data in the hope and belief that they will use this data responsibly to help, if not them, then future generations like them. From blood sample, to questionnaire, to brain donation, it is vital to treat each donation with the respect it deserves and maximise the

value that can be derived from it by accurately recording and sharing the data with the wider scientific community. Existing barriers to the timely release of data are largely cultural and, therefore, with enough leverage and support from within the scientific community, can be overcome, opening the way for essential progress in the pursuit of disease modifying therapies.

Acronyms

Aβ β-amyloid. 15, 16, 18, 30, 31

AC anterior cingulate. 105, 107, 111, 119–121

ADNI Alzheimer's Disease Neuroimaging Initiative. 44, 79, 144

APP amyloid precursor protein. 16, 18, 82

AT anterior temporal. 104, 107, 111, 115, 119, 121

AUC area under the receiver-operator characteristic curve. 107, 108, 115, 121, 147, 148, 154

bvFTD behavioural variant frontotemporal dementia. 23, 65, 67, 69, 71

C9ORF72 chromosome 9 open reading frame 72. 22, 82

CAA cerebral amyloid angiopathy. 16, 25, 59, 64

CBD corticobasal degeneration. 27, 67, 75

CBS corticobasal syndrome. 23, 65, 67, 72

CI confidence interval. 91, 93, 95

CSF cerebrospinal fluid. 30, 45, 47, 61, 78, 85, 86, 105, 122, 127

DARTEL Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra. 48, 49, 86, 87, 105, 127

DF degrees-of-freedom. 43, 47, 48, 146

DICOM Digital Imaging and Communications in Medicine. 84, 168

DLB dementia with Lewy bodies. 15, 19, 20, 22, 27, 32, 33, 72, 75, 82, 90, 92, 93, 95, 99, 101, 104, 107, 108, 115, 121, 122, 125, 128, 132, 134, 139, 141, 143, 147, 148, 150, 154, 163, 164, 177, 189

EOAD early-onset Alzheimer's disease. 125, 128, 138, 164

FI fronto insula. 105, 107, 110

FLAIR fluid attenuation inversion recovery. 25, 36, 59, 61–63, 72, 84

FTLD frontotemporal lobar degeneration. 15, 20, 22, 23, 31, 33, 35, 36, 39, 65, 69, 79, 82, 83, 90, 93, 95, 99, 101, 102, 104, 107, 108, 111, 115, 121, 122, 125, 128, 132–134, 139–141, 143, 147, 148, 150, 154, 163, 164, 177

FUS fused in sarcoma protein. 20, 22–24, 82, 83, 90, 92, 93, 99

FWE family-wise error rate. 52, 128, 132–134, 138

FWHM full-width at half-maximum. 86

GRN progranulin. 22, 75, 82

ICBMI152 International Consortium for Brain Mapping 152. 44, 86, 87

ICC intraclass correlation coefficient. 106, 110, 111, 120

INU intensity non-uniformity. 42, 47, 84, 85

LOAD Late-onset Alzheimer's disease. 125, 128, 138

MAPT microtubule associated protein tau. 22, 72, 82

MIDAS Medical Information Display and Analysis System. 85, 90

MMSE Mini-Mental State Examination. 93, 108, 128, 148

MND motor neuron disease. 22–24, 39, 71, 82

MTA medial temporal lobe atrophy. 104, 107, 110, 111, 115, 119–121, 177

MTL medial temporal lobe. 178

NFTs neurofibrillary tangles. 15, 16, 22, 71

NIA-AA National Institute on Ageing and the Alzheimer's Association. 15, 16, 18

OF orbito-frontal. 105, 107, 110, 111, 115, 120, 121

PA posterior atrophy. 104, 107, 110, 115, 119–121

PET positron emission tomography. 27, 30, 31, 78, 140

PNFA progressive non-fluent aphasia. 65, 67

PSEN1 Presenilin 1. 16, 18, 82

PSEN2 Presenilin 2. 16, 18

PSP progressive supranuclear palsy. 23, 27, 67, 75

PVS perivascular spaces. 63

SPECT single photon emission computed tomography. 27, 78

SPM statistical parametric mapping. 46, 47, 49, 50, 86, 87, 125, 127, 128, 140

STRIVE STandards for ReportIng Vascular changes on nEuroimaging. 36, 37, 60

SVC support vector classifier. 107, 108, 115, 123, 143, 147, 148, 154, 163, 165, 166, 177

SVD small vessel disease. 59, 62, 63

SVM support vector machine. 53–55, 88

TDP43 transactive response DNA binding protein 43 kDa. 16, 20, 22–24, 67, 69, 71, 79, 82, 83, 90, 92, 93, 95, 99, 102, 104, 111, 125, 128, 132–134, 140, 141, 143, 147, 148, 150, 154, 163, 164

TIV total intracranial volume. 51, 85, 86, 107, 108, 147, 150, 154, 163, 165

VBM voxel-based morphometry. 46–48, 51, 52, 84, 86, 105, 111, 124, 127, 138

Bibliography

1. Gaugler, J. E. *et al.* Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the NACC-UDS database. *BMC geriatrics* **13,** 137 (2013).
2. Schott, J. M. & Warren, J. D. Alzheimer's disease: mimics and chameleons. *Practical Neurology* **12,** 358–366 (Dec. 2012).
3. Beach, T. G., Monsell, S. E., Phillips, L. E. & Kukull, W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of Neuropathology and Experimental Neurology* **71,** 266–273 (Apr. 2012).
4. Xie, J., Brayne, C., Matthews, F. E. & the Medical Research Council Cognitive Function and Ageing Study collaborators. Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up. *BMJ* **336,** 258–262 (Feb. 2, 2008).
5. World Health Organisation. *WHO Dementia Fact Sheet 2015* Fact Sheet 362 (Mar. 2015).
6. Matthews, F. E. *et al.* A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *The Lancet* **382,** 1405–1412 (Oct. 2013).
7. Alzheimer's Society. *Dementia 2014: Opportunity for change* (Sept. 2014).
8. UK Department of Health. *Prime minister's challenge on dementia - Delivering major improvements in dementia care and research by 2015* (Mar. 2012).
9. Direction générale de l'action sociale, Direction de la sécurité sociale, Direction de l'hospitalisation et de l'organisation des soins. *Plan Alzheimer et maladies apparentées 2008-2012* (Feb. 2008).
10. US Department of Health and Human Services. *National plan to address Alzheimer's disease* (May 2012).
11. UK Department of Health. *Living well with dementia: a National Dementia Strategy* (Feb. 2009).

12. Hyman, B. T. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **8,** 1–13 (Jan. 2012).
13. Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science (New York, N.Y.)* **256,** 184–185 (Apr. 10, 1992).
14. Sambamurti, K. *et al.* A partial failure of membrane protein turnover may cause Alzheimer's disease: a new hypothesis. *Current Alzheimer Research* **3,** 81–90 (Feb. 2006).
15. Shen, J. & Kelleher, R. J. The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 403–409 (Jan. 9, 2007).
16. Small, S. A. & Duff, K. Linking Abeta and tau in late-onset Alzheimer's disease: a dual pathway hypothesis. *Neuron* **60,** 534–542 (Nov. 26, 2008).
17. Hardy, J. The amyloid hypothesis for Alzheimer's disease: a critical reappraisal. *Journal of Neurochemistry* **110,** 1129–1134 (Aug. 2009).
18. Buée, L., Bussière, T., Buée-Scherrer, V., Delacourte, A. & Hof, P. R. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Research. Brain Research Reviews* **33,** 95–130 (Aug. 2000).
19. Morris, M. *et al.* Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nature Neuroscience* **18,** 1183–1189 (Aug. 2015).
20. Tarasoff-Conway, J. M. *et al.* Clearance systems in the brain-implications for Alzheimer disease. *Nature Reviews. Neurology* (July 21, 2015).
21. Montine, T. J. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathologica* **123,** 1–11 (Jan. 2012).
22. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58,** 1791–1800 (June 25, 2002).
23. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica* **82,** 239–259 (1991).

24. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica* **112,** 389–404 (Oct. 2006).

25. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41,** 479–486 (Apr. 1991).

26. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* **65,** 1863–1872 (Dec. 2005).

27. Schneider, J. A., Arvanitakis, Z., Bang, W. & Bennett, D. A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* **69,** 2197–2204 (Dec. 11, 2007).

28. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349,** 704–706 (Feb. 21, 1991).

29. Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375,** 754–760 (June 29, 1995).

30. Rogaev, E. I. *et al.* Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376,** 775–778 (Aug. 31, 1995).

31. Bertram, L. & Tanzi, R. E. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Reviews. Neuroscience* **9,** 768–778 (Oct. 2008).

32. Saunders, A. M. *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43,** 1467–1472 (Aug. 1993).

33. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *The New England Journal of Medicine* **368,** 117–127 (Jan. 10, 2013).

34. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45,** 1452–1458 (Oct. 27, 2013).

35. Guerreiro, R. & Hardy, J. Genetics of Alzheimer's disease. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics* **11,** 732–737 (Oct. 2014).

36. Sperling, R. A. *et al.* Toward defining the preclinical stages of Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7,** 280–292 (May 2011).

37. Dubois, B. *et al.* Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet. Neurology* **13,** 614–629 (June 2014).

38. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7,** 270–279 (May 2011).

39. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7,** 263–269 (May 2011).

40. Irwin, D. J., Lee, V. M.-Y. & Trojanowski, J. Q. Parkinson's disease dementia: convergence of -synuclein, tau and amyloid- pathologies. *Nature Reviews Neuroscience* **14,** 626–636 (July 31, 2013).

41. Samuel, W., Galasko, D., Masliah, E. & Hansen, L. A. Neocortical lewy body counts correlate with dementia in the Lewy body variant of Alzheimer's disease. *Journal of Neuropathology and Experimental Neurology* **55,** 44–52 (Jan. 1996).

42. Apaydin, H., Ahlskog, J. E., Parisi, J. E., Boeve, B. F. & Dickson, D. W. Parkinson disease neuropathology: later-developing dementia and loss of the levodopa response. *Archives of Neurology* **59,** 102–112 (Jan. 2002).

43. Colosimo, C., Hughes, A. J., Kilford, L. & Lees, A. J. Lewy body cortical involvement may not always predict dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry* **74,** 852–856 (July 2003).

44. Bogaerts, V. *et al.* A novel locus for dementia with Lewy bodies: a clinically and genetically heterogeneous disorder. *Brain: A Journal of Neurology* **130,** 2277–2291 (Pt 9 Sept. 2007).

45. Clark, L. N. *et al.* Association of glucocerebrosidase mutations with dementia with lewy bodies. *Archives of Neurology* **66,** 578–583 (May 2009).

46. Bras, J. *et al.* Genetic analysis implicates APOE, SNCA and suggests lysosomal dysfunction in the etiology of dementia with Lewy bodies. *Human Molecular Genetics* **23,** 6139–6146 (Dec. 1, 2014).

47. Mackenzie, I. R. A. & Rademakers, R. The molecular genetics and neuropathology of frontotemporal lobar degeneration: recent developments. *Neurogenetics* **8,** 237–248 (Nov. 2007).

48. Mackenzie, I. R. A. *et al.* Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. *Acta Neuropathologica* **119,** 1–4 (Jan. 2010).

49. Cairns, N. J. *et al.* Neuropathologic diagnostic and nosologic criteria for frontotemporal lobar degeneration: consensus of the Consortium for Frontotemporal Lobar Degeneration. *Acta Neuropathologica* **114,** 5–22 (July 2007).

50. Rademakers, R., Neumann, M. & Mackenzie, I. R. Advances in understanding the molecular basis of frontotemporal dementia. *Nature Reviews Neurology* (June 26, 2012).

51. Lee, E. B., Lee, V. M.-Y. & Trojanowski, J. Q. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nature Reviews Neuroscience* (Nov. 30, 2011).

52. Lee, S. E. *et al.* Clinicopathological correlations in corticobasal degeneration. *Ann Neurol* **70,** 327–340 (Aug. 2011).

53. Amador-Ortiz, C. *et al.* TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Annals of Neurology* **61,** 435–445 (May 2007).

54. Nakashima-Yasuda, H. *et al.* Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathologica* **114,** 221–229 (Sept. 2007).

55. Mackenzie, I. R. A. *et al.* A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathologica* **122,** 111–113 (July 2011).

56. Munoz, D. G. *et al.* FUS pathology in basophilic inclusion body disease. *Acta Neuropathologica* **118,** 617–627 (Nov. 2009).

57. Lashley, T., Rohrer, J. D., Mead, S. & Revesz, T. Review: An update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations: Fron-

totemporal lobar degeneration, a review. *Neuropathology and Applied Neurobiology,* n/a–n/a (Aug. 2015).

58. Rohrer, J. D. *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* **73,** 1451–1456 (Nov. 3, 2009).
59. Rascovsky, K. *et al.* Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134,** 2456–2477 (Pt 9 Sept. 2011).
60. Gorno-Tempini, M. L. *et al.* Classification of primary progressive aphasia and its variants. *Neurology* **76,** 1006–1014 (Mar. 2011).
61. Warren, J. D., Rohrer, J. D. & Rossor, M. N. Frontotemporal dementia. *BMJ* **347,** f4827–f4827 (aug12 3 Aug. 20, 2013).
62. Armstrong, M. J. *et al.* Criteria for the diagnosis of corticobasal degeneration. *Neurology* **80,** 496–503 (Jan. 2013).
63. Litvan, I. *et al.* Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology* **47,** 1–9 (July 1996).
64. Devenney, E., Vucic, S., Hodges, J. R. & Kiernan, M. C. Motor neuron disease-frontotemporal dementia: a clinical continuum. *Expert Review of Neurotherapeutics* **15,** 509–522 (May 2015).
65. National Collaborating Centre for Mental Health (UK). *Dementia: The NICE-SCIE Guideline on Supporting People with Dementia and Their Carers in Health and Social Care (National Clinical Practice Guideline)* (British Psychological Society and RCPsych Publications, 2006).
66. Cooper, S. & Greene, J. D. W. The clinical assessment of the patient with early dementia. *Journal of Neurology, Neurosurgery, and Psychiatry* **76 Suppl 5,** v15–24 (Dec. 2005).
67. Rossor, M. N., Fox, N. C., Mummery, C. J., Schott, J. M. & Warren, J. D. The diagnosis of young-onset dementia. *Lancet Neurol* **9,** 793–806 (Aug. 2010).
68. Jack, C. R. *et al.* The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* **27,** 685–691 (Apr. 2008).
69. Barkof, F., Fox, N. C., Bastos-Leite, A. J. & Scheltens, P. *Neuroimaging in Dementia* (Springer Berlin Heidelberg, 2011).

70. Cheng, A.-L. *et al.* Susceptibility-weighted imaging is more reliable than T2*-weighted gradient-recalled echo MRI for detecting microbleeds. *Stroke; a Journal of Cerebral Circulation* **44,** 2782–2786 (Oct. 2013).

71. Greenberg, S. M. *et al.* Cerebral microbleeds: a guide to detection and interpretation. *The Lancet Neurology* **8,** 165–174 (Feb. 2009).

72. Gass, A., Ay, H., Szabo, K. & Koroshetz, W. J. Diffusion-weighted MRI for the small stuff: the details of acute cerebral ischaemia. *The Lancet Neurology* **3,** 39–45 (Jan. 2004).

73. Kallenberg, K. *et al.* Creutzfeldt-Jakob disease: comparative analysis of MR imaging sequences. *AJNR Am J Neuroradiol* **27,** 1459–1462 (Aug. 2006).

74. Kapucu, O. L. *et al.* EANM procedure guideline for brain perfusion SPECT using 99mTc-labelled radiopharmaceuticals, version 2. *European Journal of Nuclear Medicine and Molecular Imaging* **36,** 2093–2102 (Dec. 2009).

75. O'Brien, J. T. *et al.* 18F-FDG PET and perfusion SPECT in the diagnosis of Alzheimer and Lewy body dementias. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **55,** 1959–1965 (Dec. 2014).

76. Eshuis, S. A. *et al.* Direct comparison of FP-CIT SPECT and F-DOPA PET in patients with Parkinson's disease and healthy controls. *European Journal of Nuclear Medicine and Molecular Imaging* **36,** 454–462 (Mar. 2009).

77. Lee, C. S. *et al.* In vivo positron emission tomographic evidence for compensatory changes in presynaptic dopaminergic nerve terminals in Parkinson's disease. *Annals of Neurology* **47,** 493–503 (Apr. 2000).

78. Koeppe, R. A. *et al.* 11C-DTBZ and 18F-FDG PET measures in differentiating dementias. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **46,** 936–944 (June 2005).

79. Morgan, S. *et al.* Differentiation of frontotemporal dementia from dementia with Lewy bodies using FP-CIT SPECT. *Journal of Neurology, Neurosurgery & Psychiatry* **83,** 1063–1070 (Nov. 1, 2012).

80. Villemagne, V. L. *et al.* Amyloid imaging with (18)F-florbetaben in Alzheimer disease and other dementias. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **52,** 1210–1217 (Aug. 2011).

81. Blennow, K., Hampel, H., Weiner, M. & Zetterberg, H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews. Neurology* **6,** 131–144 (Mar. 2010).

82. Hansson, O. *et al.* Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *The Lancet. Neurology* **5,** 228–234 (Mar. 2006).

83. Mattsson, N. *et al.* CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* **302,** 385–393 (July 2009).

84. Bartlett, J. W. *et al.* Determining cut-points for Alzheimer's disease biomarkers: statistical issues, methods and challenges. *Biomarkers in Medicine* **6,** 391–400 (Aug. 2012).

85. Mason, N. S., Mathis, C. A. & Klunk, W. E. Positron emission tomography radioligands for in vivo imaging of Aβ{} plaques: PET radioligands for in vivo imaging of Aβ{} plaques. *Journal of Labelled Compounds and Radiopharmaceuticals* **56,** 89–95 (Mar. 2013).

86. Palmqvist, S. *et al.* Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid -amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA neurology* **71,** 1282–1289 (Oct. 2014).

87. Klunk, W. E. *et al.* The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **11,** pages (Jan. 2015).

88. Clark, C. M. *et al.* Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-βplaques: a prospective cohort study. *Lancet Neurol* **11,** 669–678 (Aug. 2012).

89. Wong, D. F. *et al.* In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (florbetapir [corrected] F 18). *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **51,** 913–920 (June 2010).

90. Nelissen, N. *et al.* Phase 1 study of the Pittsburgh compound B derivative 18F-flutemetamol in healthy volunteers and patients with probable Alzheimer disease. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **50,** 1251–1259 (Aug. 2009).

91. Rowe, C. C. *et al.* Imaging of amyloid beta in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. *The Lancet. Neurology* **7,** 129–135 (Feb. 2008).

92. Rowe, C. C. *et al.* Head-to-head comparison of 11C-PiB and 18F-AZD4694 (NAV4694) for -amyloid imaging in aging and dementia. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **54,** 880–886 (June 2013).

93. Landau, S. M. *et al.* Amyloid PET imaging in Alzheimers disease: a comparison of three radiotracers. *European Journal of Nuclear Medicine and Molecular Imaging* **41,** 1398–1407 (July 2014).

94. Villemagne, V. L. & Okamura, N. In vivo tau imaging: Obstacles and progress. *Alzheimer's & Dementia* **10,** S254–S264 (June 2014).

95. Villemagne, V. L., Fodero-Tavoletti, M. T., Masters, C. L. & Rowe, C. C. Tau imaging: early progress and future directions. *The Lancet Neurology* **14,** 114–124 (Jan. 2015).

96. Tolboom, N. *et al.* Detection of Alzheimer pathology in vivo using both 11C-PIB and 18F-FDDNP PET. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* **50,** 191–197 (Feb. 2009).

97. Thompson, P. W. *et al.* Interaction of the amyloid imaging tracer FDDNP with hallmark Alzheimer's disease pathologies. *Journal of Neurochemistry* **109,** 623–630 (Apr. 2009).

98. Jack, C. R. *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* **12,** 207–216 (Feb. 2013).

99. Jack, C. R. *et al.* Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* **49,** 786–794 (Sept. 1997).

100. Fox, N. C., Warrington, E. K. & Rossor, M. N. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease. *The Lancet* **353,** 2125 (June 1999).

101. Cash, D. M., Rohrer, J. D., Ryan, N. S., Ourselin, S. & Fox, N. C. Imaging endpoints for clinical trials in Alzheimers disease. *Alzheimer's Research & Therapy* **6,** 87 (2014).

102. Fox, N. C. & Freeborough, P. A. Brain atrophy progression measured from registered serial MRI: Validation and application to alzheimer's disease. *Journal of Magnetic Resonance Imaging* **7,** 1069–1075 (Nov. 1997).

103. Weiner, M. W. *et al.* Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimer's & Dementia* **11,** 865–884 (July 2015).

104. Fox, N. C. *et al.* Effects of A immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology* **64,** 1563–1572 (May 10, 2005).

105. Schott, J. M. *et al.* Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology* **65,** 119–124 (July 12, 2005).

106. Gunning-Dixon, F. M., Brickman, A. M., Cheng, J. C. & Alexopoulos, G. S. Aging of cerebral white matter: a review of MRI findings. *International Journal of Geriatric Psychiatry* **24,** 109–117 (Feb. 2009).

107. Coffey, C. E. *et al.* Quantitative cerebral anatomy of the aging human brain: a cross-sectional study using magnetic resonance imaging. *Neurology* **42,** 527–536 (Mar. 1992).

108. Good, C. *et al.* A voxel-based morphometric study of ageing in 465 normal adult human brains, II_5_1–II_5_16 (2002).

109. Likeman, M. *et al.* Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. *Arch Neurol* **62,** 1410–1415 (Sept. 2005).

110. Rohrer, J. D. *et al.* Clinical and neuroanatomical signatures of tissue pathology in frontotemporal lobar degeneration. *Brain* **134,** 2565–2581 (Pt 9 Sept. 2011).

111. Lehmann, M. *et al.* Posterior cerebral atrophy in the absence of medial temporal lobe atrophy in pathologically-confirmed Alzheimer's disease. *Neurobiol Aging* **33,** 627.e1–627.e12 (Mar. 2012).

112. Chan, D. *et al.* Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol* **49,** 433–442 (Apr. 2001).

113. Schmidt, R. *et al.* Heterogeneity in age-related white matter changes. *Acta Neuropathologica* **122,** 171–185 (Aug. 2011).

114. Raz, L., Knoefel, J. & Bhaskar, K. The neuropathology and cerebrovascular mechanisms of dementia. *Journal of Cerebral Blood Flow & Metabolism* (July 15, 2015).

115. Wardlaw, J. M. *et al.* Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol* **12,** 822–838 (Aug. 2013).

116. Fazekas, F., Chawluk, J. B., Alavi, A., Hurtig, H. I. & Zimmerman, R. A. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am J Roentgenol* **149,** 351–356 (Aug. 1987).

117. Manolio, T. A. *et al.* Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. *Stroke; a Journal of Cerebral Circulation* **25,** 318–327 (Feb. 1994).

118. Scheltens, P. *et al.* A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *Journal of the Neurological Sciences* **114,** 7–12 (Jan. 1993).

119. Wahlund, L. O. *et al.* A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* **32,** 1318–1322 (June 2001).

120. Harper, L., Barkhof, F., Fox, N. C. & Schott, J. M. Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales. *Journal of Neurology, Neurosurgery, and Psychiatry* (Apr. 14, 2015).

121. Pasquier, F. *et al.* Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *Eur Neurol* **36,** 268–272 (1996).

122. Henneman, W. J. P. *et al.* MRI biomarkers of vascular damage and atrophy predicting mortality in a memory clinic population. *Stroke* **40,** 492–498 (Feb. 2009).

123. Scheltens, P., Pasquier, F., Weerts, J. G., Barkhof, F. & Leys, D. Qualitative assessment of cerebral atrophy on MRI: inter- and intra-observer reproducibility in dementia and normal aging. *Eur Neurol* **37,** 95–99 (1997).

124. Broe, M. *et al.* Staging disease severity in pathologically confirmed cases of frontotemporal dementia. *Neurology* **60,** 1005–1011 (Mar. 2003).

125. Davies, R. R. *et al.* Progression in frontotemporal dementia: identifying a benign behavioral variant by magnetic resonance imaging. *Arch Neurol* **63,** 1627–1631 (Nov. 2006).

126. Kipps, C. M. *et al.* Clinical significance of lobar atrophy in frontotemporal dementia: application of an MRI visual rating scale. *Dement Geriatr Cogn Disord* **23,** 334–342 (2007).

127. Ambikairajah, A. *et al.* A visual MRI atrophy rating scale for the amyotrophic lateral sclerosis-frontotemporal dementia continuum. *Amyotroph Lateral Scler Frontotemporal Degener* (Feb. 2014).

128. De Leon, M. J., George, A. E., Stylopoulos, L. A., Smith, G. & Miller, D. C. Early marker for Alzheimer's disease: the atrophic hippocampus. *Lancet* **2,** 672–673 (Sept. 1989).

129. Scheltens, P. *et al.* Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* **55,** 967–972 (Oct. 1992).

130. Galton, C. J. *et al.* Temporal lobe rating scale: application to Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* **70,** 165–173 (Feb. 2001).

131. Urs, R. *et al.* Visual rating system for assessing magnetic resonance images: a tool in the diagnosis of mild cognitive impairment and Alzheimer disease. *J Comput Assist Tomogr* **33,** 73–78 (2009).

132. Duara, R. *et al.* Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. *Neurology* **71,** 1986–1992 (Dec. 2008).

133. Kim, G. *et al.* T1-Axial Medial Temporal Atrophy Visual Rating: A Comparable Study with Schelten's T1-Coronal Visual Rating. *Dement Neurocognitive Disord* **8,** 37–44 (2009).

134. Scheltens, P., Launer, L. J., Barkhof, F., Weinstein, H. C. & van Gool, W. A. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol* **242,** 557–560 (Sept. 1995).

135. Cavallin, L. *et al.* Overtime reliability of medial temporal lobe atrophy rating in a clinical setting. *Acta Radiol* **53,** 318–323 (Apr. 2012).

136. Barkhof, F. *et al.* The significance of medial temporal lobe atrophy: a postmortem MRI study in the very old. *Neurology* **69,** 1521–1527 (Oct. 2007).

137. Ferreira, D. *et al.* Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment. *Journal of Internal Medicine* (Mar. 9, 2015).

138. Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6,** 734–746 (Aug. 2007).

139. Koedam, E. L. G. E. *et al.* Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol* **21,** 2618–2625 (Dec. 2011).

140. Belaroussi, B., Milles, J., Carme, S., Zhu, Y. M. & Benoit-Cattin, H. Intensity non-uniformity correction in MRI: Existing methods and their validation. *Medical Image Analysis* **10,** 234–246 (Apr. 2006).

141. Simmons, A., Tofts, P. S., Barker, G. J. & Arridge, S. R. Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine* **32,** 121–128 (July 1994).

142. Vovk, U., Pernus, F. & Likar, B. A Review of Methods for Correction of Intensity Inhomogeneity in MRI. *IEEE Transactions on Medical Imaging* **26,** 405–421 (Mar. 2007).

143. Arnold, J. B. *et al.* Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage* **13,** 931–943 (May 2001).

144. O'Donovan, J. *et al.* Does posterior cortical atrophy on MRI discriminate between Alzheimer's disease, dementia with Lewy bodies, and normal aging? *Int Psychogeriatr* **25,** 111–119 (Jan. 2013).

145. Davies, R. R. *et al.* Development of an MRI rating scale for multiple brain regions: comparison with volumetrics and with voxel-based morphometry. *Neuroradiology* **51,** 491–503 (Aug. 2009).

146. Chow, T. W. *et al.* Visual rating versus volumetry to detect frontotemporal dementia. *Dement Geriatr Cogn Disord* **31,** 371–378 (2011).

147. De Leon, M. J. *et al.* Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. *Neurobiol Aging* **18,** 1–11 (1997).

148. Kaneko, T. *et al.* New visual rating system for medial temporal lobe atrophy: a simple diagnostic tool for routine examinations. *Psychogeriatrics* **12,** 88–92 (June 2012).

149. Hill, D. L. G., Batchelor, P. G., Holden, M. & Hawkes, D. J. Medical image registration. *Physics in Medicine and Biology* **46,** R1–R45 (Mar. 1, 2001).

150. Geuze, E., Vermetten, E. & Bremner, J. D. MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed. *Molecular Psychiatry* **10,** 147–159 (Feb. 2005).

151. Konrad, C. *et al.* Defining the human hippocampus in cerebral magnetic resonance images–an overview of current segmentation protocols. *NeuroImage* **47,** 1185–1195 (Oct. 1, 2009).

152. Boccardi, M. *et al.* Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. *Alzheimer's & Dementia* **11,** 126–138 (Feb. 2015).

153. Despotovi, I., Goossens, B. & Philips, W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Computational and Mathematical Methods in Medicine* **2015,** 1–23 (2015).

154. Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* **47,** S102 (July 2009).

155. Talairach, J. & Tournoux, P. *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging* 122 pp. (Georg Thieme, Stuttgart ; New York, 1988).

156. Brett, M., Christoff, K., Cusack, R. & Lancaster, J. Using the Talairach atlas with the MNI template. *Neuroimage* **13,** 85–85 (June 2001).

157. Evans, A. C., Janke, A. L., Collins, D. L. & Baillet, S. Brain templates and atlases. *NeuroImage* **62,** 911–922 (Aug. 15, 2012).

158. Battaglini, M., Smith, S. M., Brogi, S. & De Stefano, N. Enhanced brain extraction improves the accuracy of brain atrophy estimation. *NeuroImage* **40,** 583–589 (Apr. 1, 2008).

159. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J. & Bach Cuadra, M. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine* **104,** e158–e177 (Dec. 2011).

160. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9,** 179–194 (Feb. 1999).

161. Ashburner, J. & Friston, K. J. Voxel-Based MorphometryThe Methods. *NeuroImage* **11,** 805–821 (June 2000).

162. Ashburner, J. & Friston, K. J. Unified segmentation. *Neuroimage* **26,** 839–851 (July 2005).

163. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38,** 95–113 (Oct. 2007).

164. Ashburner, J. & Friston, K. J. Computing average shaped tissue probability templates. *NeuroImage* **45,** 333–341 (Apr. 1, 2009).

165. Ridgway, G. R. *et al.* Issues with threshold masking in voxel-based morphometry of atrophied brains. *Neuroimage* **44,** 99–111 (Jan. 2009).

166. Friston, K. J. *et al.* Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **2,** 189–210 (1994).

167. Barnes, J. *et al.* Head size, age and gender adjustment in MRI studies: a necessary nuisance? *NeuroImage* **53,** 1244–1255 (Dec. 2010).

168. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15,** 870–878 (Apr. 2002).

169. Nichols, T. & Hayasaka, S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research* **12,** 419–446 (Oct. 2003).

170. Brett, M., Penny, W. & Stefan, K. in *Human Brain Function* 2nd ed. (Mar. 2003).

171. Wernick, M., Yang, Y., Brankov, J., Yourganov, G. & Strother, S. Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine* **27,** 25–38 (July 2010).

172. Lemm, S., Blankertz, B., Dickhaus, T. & Müller, K.-R. Introduction to machine learning for brain imaging. *NeuroImage* **56,** 387–399 (May 2011).

173. Harper, L., Barkhof, F., Scheltens, P., Schott, J. M. & Fox, N. C. An algorithmic approach to structural imaging in dementia. *J Neurol Neurosurg Psychiatry* **85,** 692–698 (June 2014).

174. Gorelick, P. B. *et al.* Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke* **42,** 2672–2713 (Sept. 2011).

175. Knopman, D. S. & Roberts, R. Vascular Risk Factors: Imaging and Neuropathologic Correlates. *Journal of Alzheimer's Disease,* 699–709 (2010).

176. Thal, D. R., Grinberg, L. T. & Attems, J. Vascular dementia: Different forms of vessel disorders contribute to the development of dementia in the elderly brain. *Experimental Gerontology* **47,** 816–824 (Nov. 2012).

177. Román, G. C. *et al.* Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* **43,** 250–260 (Feb. 1993).

178. Van Straaten, E. C. W. *et al.* Operational definitions for the NINDS-AIREN criteria for vascular dementia: an interobserver study. *Stroke* **34,** 1907–1912 (Aug. 2003).

179. Savoiardo, M. The vascular territories of the carotid and vertebrobasilar systems. Diagrams based on CT studies of infarcts. *The Italian Journal of Neurological Sciences* **7,** 405–409 (July 1986).

180. Snowdon, D. A. *et al.* Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. *JAMA* **277,** 813–817 (Mar. 12, 1997).

181. Erten-Lyons, D. *et al.* Neuropathologic basis of white matter hyperintensity accumulation with advanced age. *Neurology* **81,** 977–983 (Sept. 10, 2013).

182. Dolan, H. *et al.* Atherosclerosis, dementia, and alzheimer's disease in the BLSA cohort. *Annals of Neurology,* n/a–n/a (Apr. 14, 2010).

183. Suemoto, C. K. *et al.* Atherosclerosis and dementia: a cross-sectional study with pathological analysis of the carotid arteries. *Stroke; a Journal of Cerebral Circulation* **42,** 3614–3615 (Dec. 2011).

184. Hachinski, V. *et al.* National Institute of Neurological Disorders and Stroke-Canadian Stroke Network vascular cognitive impairment harmonization standards. *Stroke; a Journal of Cerebral Circulation* **37,** 2220–2241 (Sept. 2006).

185. Bastos Leite, A. J., van Straaten, E. C. W., Scheltens, P., Lycklama, G. & Barkhof, F. Thalamic lesions in vascular dementia: low sensitivity of fluid-attenuated inversion recovery (FLAIR) imaging. *Stroke* **35,** 415–419 (Feb. 2004).

186. Koch, S., McClendon, M. S. & Bhatia, R. Imaging evolution of acute lacunar infarction: leukoariosis or lacune? *Neurology* **77,** 1091–1095 (Sept. 13, 2011).

187. Potter, G. M. *et al.* Counting cavitating lacunes underestimates the burden of lacunar infarction. *Stroke; a Journal of Cerebral Circulation* **41,** 267–272 (Feb. 2010).

188. Grau, A. J. *et al.* Risk Factors, Outcome, and Treatment in Subtypes of Ischemic Stroke: The German Stroke Data Bank. *Stroke* **32,** 2559–2566 (Nov. 1, 2001).

189. De Laat, K. F. *et al.* Gait in Elderly With Cerebral Small Vessel Disease. *Stroke* **41,** 1652–1658 (Aug. 1, 2010).

190. Vermeer, S. E. *et al.* Silent brain infarcts and the risk of dementia and cognitive decline. *The New England Journal of Medicine* **348,** 1215–1222 (Mar. 27, 2003).

191. Gouw, A. A. *et al.* Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations. *Journal of Neurology, Neurosurgery & Psychiatry* **82,** 126–135 (Feb. 1, 2011).

192. Mäntylä, R. *et al.* Variable agreement between visual rating scales for white matter hyperintensities on MRI. Comparison of 13 rating scales in a poststroke cohort. *Stroke; a Journal of Cerebral Circulation* **28,** 1614–1623 (Aug. 1997).

193. Kapeller, P. *et al.* Visual rating of age-related white matter changes on magnetic resonance imaging: scale comparison, interrater agreement, and correlations with quantitative measurements. *Stroke; a Journal of Cerebral Circulation* **34,** 441–445 (Feb. 2003).

194. Marjo Van Der Valk, J. V. *Magnetic Resonance of Myelination and Myelin Disorders* illustrated. 1084 pp. (Springer, 2005).

195. Federico, A. *et al.* Hereditary cerebral small vessel diseases: a review. *J Neurol Sci* **322,** 25–30 (Nov. 2012).

196. Choi, J. C. Genetics of cerebral small vessel disease. *Journal of Stroke* **17,** 7–16 (Jan. 2015).

197. Marín-Padilla, M. & Knopman, D. S. Developmental Aspects of the Intracerebral Microvasculature and Perivascular Spaces: Insights into Brain Response to Late-Life Diseases. *Journal of Neuropathology & Experimental Neurology* **70,** 1060–1069 (Dec. 2011).

198. Charidimou, A. *et al.* White matter perivascular spaces: An MRI marker in pathology-proven cerebral amyloid angiopathy? *Neurology* **82,** 57–62 (Jan. 7, 2014).

199. Zhu, Y.-C. *et al.* Severity of Dilated Virchow-Robin Spaces Is Associated With Age, Blood Pressure, and MRI Markers of Small Vessel Disease: A Population-Based Study. *Stroke* **41,** 2483–2490 (Nov. 1, 2010).

200. Groeschel, S., Chong, W. K., Surtees, R. & Hanefeld, F. Virchow-Robin spaces on magnetic resonance images: normative data, their dilatation, and a review of the literature. *Neuroradiology* **48,** 745–754 (Oct. 2, 2006).

201. Potter, G. M., Chappell, F. M., Morris, Z. & Wardlaw, J. M. Cerebral Perivascular Spaces Visible on Magnetic Resonance Imaging: Development of a Qualitative Rating Scale and its Observer Reliability. *Cerebrovascular Diseases* **39,** 224–231 (Mar. 19, 2015).

202. Schrag, M. *et al.* Correlation of hypointensities in susceptibility-weighted images to tissue histology in dementia patients with cerebral amyloid angiopathy: a postmortem MRI study. *Acta Neuropathologica* **119,** 291–302 (Mar. 2010).

203. Gregoire, S. M. *et al.* Choice of echo time on GRE T2*-weighted MRI influences the classification of brain microbleeds. *Clinical Radiology* **65,** 391–394 (May 2010).

204. Charidimou, A., Krishnan, A., Werring, D. J. & Rolf Jäger, H. Cerebral microbleeds: a guide to detection and clinical relevance in different disease settings. *Neuroradiology* **55,** 655–674 (June 2013).

205. Cordonnier, C. & van der Flier, W. M. Brain microbleeds and Alzheimer's disease: innocent observation or key player? *Brain* **134,** 335–344 (Pt 2 Feb. 2011).

206. Poels, M. M. F. *et al.* Incidence of cerebral microbleeds in the general population: the Rotterdam Scan Study. *Stroke* **42,** 656–661 (Mar. 2011).

207. Seeley, W. W. *et al.* Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. *Archives of Neurology* **65,** 249–255 (Feb. 2008).

208. Lindberg, O. *et al.* Cortical morphometric subclassification of frontotemporal lobar degeneration. *AJNR Am J Neuroradiol* **30,** 1233–1239 (June 2009).

209. Hornberger, M. *et al.* Orbitofrontal dysfunction discriminates behavioral variant frontotemporal dementia from Alzheimer's disease. *Dement Geriatr Cogn Disord* **30,** 547–552 (2010).

210. Whitwell, J. L. *et al.* Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. *Brain* **132,** 2932–2946 (Nov. 1, 2009).

211. Lashley, T. *et al.* A comparative clinical, pathological, biochemical and genetic study of fused in sarcoma proteinopathies. *Brain* **134,** 2548–2564 (Sept. 1, 2011).

212. Whitwell, J. L. *et al.* Frontal asymmetry in behavioral variant frontotemporal dementia: clinicoimaging and pathogenetic correlates. *Neurobiol Aging* **34,** 636–639 (Feb. 2013).

213. Gorno-Tempini, M. L. *et al.* Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol* **55,** 335–346 (Mar. 2004).

214. Josephs, K. A. *et al.* Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain: A Journal of Neurology* **129,** 1385–1398 (Pt 6 June 2006).

215. Rohrer, J. D. *et al.* Rates of hemispheric and lobar atrophy in the language variants of frontotemporal lobar degeneration. *J Alzheimers Dis* **30,** 407–411 (2012).

216. Caso, F. *et al.* In vivo signatures of nonfluent/agrammatic primary progressive aphasia caused by FTLD pathology. *Neurology* **82,** 239–247 (Jan. 21, 2014).

217. Josephs, K. A. *et al.* Voxel-based morphometry in autopsy proven PSP and CBD. *Neurobiology of Aging* **29,** 280–289 (Feb. 2008).

218. Josephs, K. A. *et al.* Anatomical differences between CBS-corticobasal degeneration and CBS-Alzheimer's disease. *Movement Disorders* **25,** 1246–1252 (July 15, 2010).

219. Hassan, A. *et al.* Symmetric corticobasal degeneration (S-CBD). *Parkinsonism & Related Disorders* **16,** 208–214 (Mar. 2010).

220. Whitwell, J. L. *et al.* Imaging correlates of pathology in corticobasal syndrome. *Neurology* **75,** 1879–1887 (Nov. 23, 2010).

221. Rankin, K. P. *et al.* Behavioral Variant Frontotemporal Dementia with Corticobasal Degeneration Pathology: Phenotypic Comparison to bvFTD with Picks Disease. *Journal of Molecular Neuroscience* **45,** 594–608 (Nov. 2011).

222. Whitwell, J. L. *et al.* Imaging Signatures of Molecular Pathology in Behavioral Variant Frontotemporal Dementia. *Journal of Molecular Neuroscience* **45,** 372–378 (Nov. 2011).

223. Massey, L. A. *et al.* Conventional magnetic resonance imaging in confirmed progressive supranuclear palsy and multiple system atrophy: cMRI in PSP and MSA. *Movement Disorders* **27,** 1754–1762 (Dec. 2012).

224. Boeve, B. F. *et al.* Pathologic heterogeneity in clinically diagnosed corticobasal degeneration. *Neurology* **53,** 795–800 (Sept. 1999).

225. Whitwell, J. L. *et al.* Magnetic resonance imaging signatures of tissue pathology in frontotemporal dementia. *Archives of Neurology* **62,** 1402–1408 (Sept. 2005).

226. Gröschel, K. *et al.* Magnetic resonance imaging-based volumetry differentiates progressive supranuclear palsy from corticobasal degeneration. *NeuroImage* **21,** 714–724 (Feb. 2004).

227. Beck, J. *et al.* A distinct clinical, neuropsychological and radiological phenotype is associated with progranulin gene mutations in a large UK series. *Brain* **131,** 706–720 (Pt 3 Mar. 2008).

228. Rohrer, J. D. *et al.* Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations. *Neuroimage* **53,** 1070–1076 (Nov. 2010).

229. Whitwell, J. L. *et al.* Does TDP-43 type confer a distinct pattern of atrophy in frontotemporal lobar degeneration? *Neurology* **75,** 2212–2220 (Dec. 14, 2010).

230. Rohrer, J. D. & Warren, J. D. Phenotypic signatures of genetic frontotemporal dementia. *Curr Opin Neurol* **24,** 542–549 (Dec. 2011).

231. Coon, E. A., Whitwell, J. L., Parisi, J. E., Dickson, D. W. & Josephs, K. A. Right temporal variant frontotemporal dementia with motor neuron disease. *Journal of Clinical Neuroscience* **19,** 85–91 (Jan. 2012).

232. Josephs, K. A. *et al.* Corticospinal tract degeneration associated with TDP-43 type C pathology and semantic dementia. *Brain* **136,** 455–470 (Feb. 1, 2013).

233. Henry, M. L. *et al.* Neuropsychological, behavioral, and anatomical evolution in right temporal variant frontotemporal dementia: A longitudinal and post-mortem single case analysis. *Neurocase* **20,** 100–109 (Jan. 2, 2014).

234. Chan, D. *et al.* The clinical profile of right temporal lobe atrophy. *Brain* **132,** 1287–1298 (Pt 5 May 2009).

235. Rohrer, J. D. *et al.* TDP-43 subtypes are associated with distinct atrophy patterns in frontotemporal dementia. *Neurology* **75,** 2204–2211 (Dec. 14, 2010).

236. Rosen, H. J. *et al.* Patterns of brain atrophy in frontotemporal dementia and semantic dementia. *Neurology* **58,** 198–208 (Jan. 22, 2002).

237. Rosen, H. J. *et al.* Neuroanatomical correlates of behavioural disorders in dementia. *Brain* **128,** 2612–2625 (Pt 11 Nov. 2005).

238. Rohrer, J. D. *et al.* Progressive logopenic/phonological aphasia: erosion of the language network. *Neuroimage* **49,** 984–993 (Jan. 2010).

239. Hu, W. T. *et al.* Multimodal predictors for Alzheimer disease in nonfluent primary progressive aphasia. *Neurology* **75,** 595–602 (Aug. 17, 2010).

240. Whitwell, J. L. & Josephs, K. A. Neuroimaging in frontotemporal lobar degenerationpredicting molecular pathology. *Nat Rev Neurol* **8,** 131–142 (2011).

241. Josephs, K. A. *et al.* Quantitative neurofibrillary tangle density and brain volumetric MRI analyses in Alzheimers disease presenting as logopenic progressive aphasia. *Brain and Language* **127,** 127–134 (Nov. 2013).

242. Silbert, L. C. *et al.* Changes in premorbid brain volume predict Alzheimer's disease pathology. *Neurology* **61,** 487–492 (Aug. 26, 2003).

243. Whitwell, J. L. *et al.* MRI correlates of neurofibrillary tangle pathology at autopsy: a voxel-based morphometry study. *Neurology* **71,** 743–749 (Sept. 2, 2008).

244. Kantarci, K. *et al.* Focal atrophy on MRI and neuropathologic classification of dementia with Lewy bodies. *Neurology* **79,** 553–560 (Aug. 7, 2012).

245. Murray, M. E. *et al.* MRI and pathology of REM sleep behavior disorder in dementia with Lewy bodies. *Neurology* **81,** 1681–1689 (Nov. 5, 2013).

246. Kaur, B. *et al.* Association Between Neuropathology and Brain Volume in The Framingham Heart Study: *Alzheimer Disease & Associated Disorders* **28,** 219–225 (2014).

247. Nedelska, Z. *et al.* Pattern of brain atrophy rates in autopsy-confirmed dementia with Lewy bodies. *Neurobiology of Aging* **36,** 452–461 (Jan. 2015).

248. Harvey, G. T. *et al.* Magnetic resonance imaging differences between dementia with Lewy bodies and Alzheimer's disease: a pilot study. *Psychological Medicine* **29,** 181–187 (Jan. 1999).

249. Jack, C. R. *et al.* Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology* **58,** 750–757 (Mar. 12, 2002).

250. Burton, E. J. *et al.* Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: a prospective study with pathological verification of diagnosis. *Brain* **132,** 195–203 (Pt 1 Jan. 2009).

251. Burton, E. J. *et al.* Quantitative neurodegenerative pathology does not explain the degree of hippocampal atrophy on MRI in degenerative dementia: Neuropathology and hippocampal atrophy in dementia. *International Journal of Geriatric Psychiatry* **27,** 1267–1274 (Dec. 2012).

252. Ince, P., Irving, D., MacArthur, F. & Perry, R. H. Quantitative neuropathological study of Alzheimer-type pathology in the hippocampus: comparison of senile dementia of Alzheimer type, senile dementia of Lewy body type, Parkinson's disease and non-demented elderly control patients. *J Neurol Sci* **106,** 142–152 (Dec. 1991).

253. Cendes, F. Neuroimaging in investigation of patients with epilepsy. *Continuum (Minneapolis, Minn.)* **19,** 623–642 (June 2013).

254. Mahoney, C. J. *et al.* Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain* **135,** 736–750 (Mar. 1, 2012).

255. Whitwell, J. L. *et al.* Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain* **135,** 794–806 (Pt 3 Mar. 2012).

256. Rohrer, J. D. *et al.* Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis. *The Lancet Neurology* **14,** 253–262 (Mar. 2015).

257. Crutch, S. J. *et al.* Posterior cortical atrophy. *The Lancet Neurology* **11,** 170–178 (Feb. 2012).

258. Beyer, M. K., Larsen, J. P. & Aarsland, D. Gray matter atrophy in Parkinson disease with dementia and dementia with Lewy bodies. *Neurology* **69,** 747–754 (Aug. 21, 2007).

259. Watson, R., Colloby, S. J., Blamire, A. M. & OBrien, J. T. Assessment of Regional Gray Matter Loss in Dementia with Lewy Bodies: A Surface-Based MRI Analysis. *The American Journal of Geriatric Psychiatry* **23,** 38–46 (Jan. 2015).

260. Grossman, M. *et al.* Distinct antemortem profiles in patients with pathologically defined frontotemporal dementia. *Archives of Neurology* **64,** 1601–1609 (Nov. 2007).

261. Apostolova, L. G. *et al.* Cortical and Hippocampal Atrophy in Patients with Autosomal Dominant Familial Alzheimers Disease. *Dementia and Geriatric Cognitive Disorders* **32,** 118–125 (2011).

262. Benzinger, T. L. S. *et al.* Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease. *Proceedings of the National Academy of Sciences* **110,** E4502–E4509 (Nov. 19, 2013).

263. Quiroz, Y. T. *et al.* Cortical atrophy in presymptomatic Alzheimer's disease presenilin 1 mutation carriers. *Journal of Neurology, Neurosurgery & Psychiatry* **84,** 556–561 (May 1, 2013).

264. Sala-Llonch, R. *et al.* Evolving brain structural changes in PSEN1 mutation carriers. *Neurobiology of Aging* **36,** 1261–1270 (Mar. 2015).

265. Josephs, K. A. *et al.* Progressive aphasia secondary to Alzheimer disease vs FTLD pathology. *Neurology* **70,** 25–34 (Jan. 1, 2008).

266. Kantarci, K. *et al.* Multimodality imaging characteristics of dementia with Lewy bodies. *Neurobiology of Aging* **33,** 2091–2105 (Sept. 2012).

267. Koga, S. *et al.* When DLB, PD, and PSP masquerade as MSA: An autopsy study of 134 patients. *Neurology* **85,** 404–412 (Aug. 4, 2015).

268. Massey, L. A. *et al.* The midbrain to pons ratio: a simple and specific MRI sign of progressive supranuclear palsy. *Neurology* **80,** 1856–1861 (May 14, 2013).

269. Kato, N., Arai, K. & Hattori, T. Study of the rostral midbrain atrophy in progressive supranuclear palsy. *Journal of the Neurological Sciences* **210,** 57–60 (June 2003).

270. Adachi, M., Kawanami, T., Ohshima, H., Sugai, Y. & Hosoya, T. Morning glory sign: a particular MR finding in progressive supranuclear palsy. *Magnetic resonance in medical sciences: MRMS: an official journal of Japan Society of Magnetic Resonance in Medicine* **3,** 125–132 (Dec. 15, 2004).

271. Righini, A. *et al.* MR imaging of the superior profile of the midbrain: differential diagnosis between progressive supranuclear palsy and Parkinson disease. *AJNR. American journal of neuroradiology* **25,** 927–932 (July 2004).

272. Schrag, A. *et al.* Differentiation of atypical parkinsonian syndromes with routine MRI. *Neurology* **54,** 697–702 (Feb. 2000).

273. Malone, I. B. *et al.* MIRIAD: Public release of a multiple time point Alzheimer's MR imaging dataset. *NeuroImage* **70,** 33–36 (Apr. 2013).

274. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging* **17,** 87–97 (Feb. 1998).

275. Freeborough, P. A., Fox, N. C. & Kitney, R. I. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine* **53,** 15–25 (May 1997).

276. Ridgway, G., Barnes, J., Pepple, T. & Fox, N. Estimation of total intracranial volume; a comparison of methods. *Alzheimer's & Dementia* **7,** S62–S63 (July 2011).

277. Modat, M. *et al.* Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed* **98,** 278–284 (June 2010).

278. Ourselin, S., Roche, A., Subsol, G., Pennec, X. & Ayache, N. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing* **19,** 25–31 (Jan. 2001).

279. Hanke, M. *et al.* PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics* **7,** 37–53 (Mar. 2009).

280. Gouws, A., Woods, W., Millman, R., Morland, A. & Green, G. DataViewer3D: An Open-Source, Cross-Platform Multi-Modal Neuroimaging Data Visualization Tool. *Frontiers in Neuroinformatics* **3,** 9 (2009).

281. Millman, K. J. & Brett, M. Analysis of Functional Magnetic Resonance Imaging in Python. *Computing in Science & Engineering* **9,** 52–55 (2007).

282. Van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **13,** 22–30 (Mar. 2011).

283. Jones, E., Oliphant, T. & Peterson, P. *SciPy: Open source scientific tools for Python* (2001).

284. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9,** 90–95 (2007).

285. McKinney, W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference* (eds Walt, S. v. d. & Millman, J.) 51–56 (2010).

286. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

287. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33,** 159–174 (Mar. 1977).

288. Fumagalli, G. G. *et al.* 9th International Conference on Frontotemporal Dementias P.252 Development of a visual rating scale for atrophy of the anterior cingulate, insula and frontal lobes. *Am J Neurodegener Dis* **3,** 1–375 (Supplementary Issue 1 2014).

289. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* **86,** 420–428 (Mar. 1979).

290. Newcombe, R. G. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: asymptotic methods and evaluation. *Statistics in Medicine* **25,** 559–573 (Feb. 28, 2006).

291. Rakotomamonjy, A. Variable selection using SVM based criteria. *The Journal of Machine Learning Research* **3,** 1357–1370 (2003).

292. Moller, C. *et al.* Quantitative regional validation of the visual rating scale for posterior cortical atrophy. *Eur Radiol* **24,** 397–404 (Feb. 2014).

293. Förstl, H., Burns, A., Luthert, P., Cairns, N. & Levy, R. The Lewy-body variant of Alzheimer's disease. Clinical and pathological findings. *The British Journal of Psychiatry: The Journal of Mental Science* **162,** 385–392 (Mar. 1993).

294. Ahmed, R. M. *et al.* Biomarkers in dementia: clinical utility and new directions. *Journal of Neurology, Neurosurgery, and Psychiatry* **85,** 1426–1434 (Dec. 2014).

295. Ewers, M. *et al.* CSF biomarkers for the differential diagnosis of Alzheimer's disease. A large-scale international multicenter study. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* (Mar. 21, 2015).

296. Coupé, P. *et al.* Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage: Clinical* **1,** 141–152 (2012).

297. Koikkalainen, J. *et al.* Improved Classification of Alzheimer's Disease Data via Removal of Nuisance Variability. *PLoS ONE* **7** (ed Smalheiser, N. R.) e31112 (Feb. 13, 2012).

298. Whitwell, J. L. *et al.* Temporoparietal atrophy: a marker of AD pathology independent of clinical diagnosis. *Neurobiology of Aging* **32,** 1531–1541 (Sept. 2011).

299. Scahill, R. I. *et al.* Genetic Influences on Atrophy Patterns in Familial Alzheimer's Disease: A Comparison of APP and PSEN1 Mutations. *Journal of Alzheimer's Disease,* 199–212 (2013).

300. Ossenkoppele, R. *et al.* The behavioural/dysexecutive variant of Alzheimers disease: clinical, neuroimaging and pathological features. *Brain* **138,** 2732–2749 (Sept. 2015).

301. Vemuri, P. *et al.* Antemortem differential diagnosis of dementia pathology using structural MRI: Differential-STAND. *NeuroImage* **55,** 522–531 (Mar. 2011).

302. Rosenberg-Katz, K. *et al.* Gray matter atrophy distinguishes between Parkinson disease motor subtypes. *Neurology* **80,** 1476–1484 (Apr. 16, 2013).

303. Han, X. *et al.* Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage* **32,** 180–194 (Aug. 2006).

304. Barnes, J. *et al.* Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Archives of Neurology* **63,** 1434–1439 (Oct. 2006).

305. Burton, E. J. *et al.* Neuropathological correlates of volumetric MRI in autopsy-confirmed Lewy body dementia. *Neurobiology of Aging* **33,** 1228–1236 (July 2012).

306. Dawe, R. J., Bennett, D. A., Schneider, J. A. & Arfanakis, K. Neuropathologic Correlates of Hippocampal Atrophy in the Elderly: A Clinical, Pathologic, Postmortem MRI Study. *PLoS ONE* **6** (ed Pant, H.) e26286 (Oct. 17, 2011).

307. Keihaninejad, S. *et al.* Automated segmentation of the thalamus for measurement of MR diffusion indices: evaluation and application to familial Alzheimer's disease. *Alzheimer's & Dementia* **7,** S50 (July 2011).

308. Colloby, S. J., OBrien, J. T. & Taylor, J.-P. Patterns of cerebellar volume loss in dementia with Lewy bodies and Alzheimers disease: A VBM-DARTEL study. *Psychiatry Research: Neuroimaging* **223,** 187–191 (Sept. 2014).

309. Cardoso, M. J. *et al.* STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. *Med Image Anal* **17,** 671–684 (Aug. 2013).

310. Hammers, A. *et al.* Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* **19,** 224–247 (Aug. 2003).

311. Hammers, A. *et al.* Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. *Human Brain Mapping* **28,** 34–48 (Jan. 2007).

312. Marcus, D. S. *et al.* Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience* **19,** 1498–1507 (Sept. 2007).

313. Leung, K. K. *et al.* Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. *Neuroimage* **55,** 1091–1108 (Apr. 2011).

314. Warfield, S. K., Zou, K. H. & Wells, W. M. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* **23,** 903–921 (July 2004).

315. Mourão-Miranda, J. *et al.* Patient classification as an outlier detection problem: an application of the One-Class Support Vector Machine. *NeuroImage* **58,** 793–804 (Oct. 1, 2011).

316. Kloppel, S. *et al.* Automatic classification of MR scans in Alzheimer's disease. *Brain* **131,** 681–689 (Feb. 7, 2008).

317. World Dementia Council. *The World Dementia Council's Year-On Report 2014/15* 1 (2015).

318. Cooper, G. M. *The cell: a molecular approach* 2. ed. 689 pp. (ASM Press [u.a.], Washington, DC, 2000).

319. Ballatore, C., Lee, V. M.-Y. & Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nature Reviews. Neuroscience* **8,** 663–672 (Sept. 2007).

320. Duara, R. *et al.* The utility of age-specific cut-offs for visual rating of medial temporal atrophy in classifying Alzheimer's disease, MCI and cognitively normal elderly subjects. *Front Aging Neurosci* **5,** 47 (2013).

321. Koedam, E. L. G. E. *et al.* Clinical characteristics of patients with frontotemporal dementia with and without lobar atrophy on MRI. *Alzheimer Dis Assoc Disord* **24,** 242–247 (2010).

322. Westman, E. *et al.* Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS One* **6,** e22506 (2011).

323. *Bermuda principles - Policies on Release of Human Genomic Sequence Data 2003* <http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml> (2003).

324. Wellcome Trust. *Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility* (2003).

325. Rodriguez, H. *et al.* Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: the Amsterdam principles. *Journal of Proteome Research* **8,** 3689–3692 (July 2009).

326. Toronto International Data Release Workshop Authors *et al.* Prepublication data sharing. *Nature* **461,** 168–170 (Sept. 10, 2009).

327. *Dementia Research and Care* (OECD Publishing, Feb. 3, 2015).

328. Matosin, N., Frank, E., Engel, M., Lum, J. S. & Newell, K. A. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Disease Models & Mechanisms* **7,** 171–173 (Feb. 2014).

329. Journals unite for reproducibility. *Nature* **515,** 7–7 (Nov. 5, 2014).

330. Economic and Social Research Council. *ESRC Research Data Policy* <`http://www.esrc.ac.uk/funding/guidance-for-grant-holders/research-data-policy/`> (2015).

331. Toombs, J. *et al.* Identification of an important potential confound in CSF AD studies: aliquot volume. *Clinical chemistry and laboratory medicine: CCLM / FESCC* **51,** 2311–2317 (Dec. 2013).